

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



## POWER PLAY

Energy-producing  
mitochondria came  
late to eukaryote-like  
cells **PAGES 39 & 101**

**NATURAL GAS**

### FRACKING DISAPPOINTMENT

Europe falling to match  
US shale-gas bonanza

**PAGE 22**

**POLICY**

### FUKUSHIMA FIVE YEARS ON

What did Japan learn  
about energy and security?

**PAGE 29**

**NUTRITION**

### DIETARY FAT A CANCER RISK

Pro-obesity diet promotes  
'stemness' in intestinal cells

**PAGES 42 & 53**

**NATURE.COM/NATURE**

3 March 2016 \$10

Vol 531, No 7592





# Unintended consequences

**After the introduction of a clumsily worded new rule, the UK government should move quickly to reassure scientists that they can continue to advise policymakers.**

02 March 2016

Only a fool ignores well informed advice. And only a very foolish government demands not to receive it in the first place.

But that is what the British government is in danger of doing. Last month the Cabinet Office — the ministry that supports the Prime Minister in running the government — introduced a new condition attached to government grants.

A new rule warns that money from any grant, either issued direct from departments or through third parties, cannot be used to “support activity intended to influence or attempt to influence Parliament, government or political parties ... or attempting to influence legislative or regulatory action”.

## Related stories

- Confusion reigns as UK scientists face government ‘gagging’ clause
- Canadian election brings hope for science
- Communication breakdown

Despite increasing concern from academics, the Department for Business, Innovation and Skills, responsible for billions of pounds of research funding, could not say as *Nature* went to press whether the rule would apply to science grants and university funding. The research councils and the Higher Education Funding Council for England — which actually parcel up the money for institutions and academics — are equally in the dark. All of which leaves scientists fearing that they are about to be muzzled.

**“The UK case seems to be more cock-up than conspiracy.”**

This situation is not as shocking as in Canada, where a previous government deliberately set out to gag its researchers. Instead, the UK case seems to be more cock-up than conspiracy. No official seems to have thought through what it might mean to stop anyone who receives government money saying anything of substance to

government. The Cabinet Office says that the clause was introduced to stop bodies that rely on government funds from lobbying government for more funding.

What could be lost if this clause is implemented fully is unclear. The specifics of how it will work have not been set out in great detail. But it could cover some of what government-funded scientists already do. A group of cross-party politicians in the House of Lords, for example, is conducting an inquiry into what impact the result of the United Kingdom’s pending referendum, on whether to stay in or quit the European Union, would have on science. Among those giving evidence to the Lords are seven research



institutions that are either government-owned or receive substantial government grants, including the Met Office weather agency, animal-health centre the Pirbright Institute in Surrey and the plant experts at the John Innes Centre in Norwich. Then there are nine universities or university centres, plus individual professors.

All the evidence culled from this wealth of expertise could be jeopardized by a heavy-handed implementation of this clause — for what is the point of evidence that has no influence? Even if these groups did still give evidence, some of it would have to be watered down or heavily qualified. Academic input has enlightened discussions of climate change, pollinator declines, biomedical ethics and many other issues of crucial importance to the future of the United Kingdom and the wider world.

The clause does not even limit itself to activity that tries to influence the United Kingdom; it merely forbids “attempting to influence legislative or regulatory action”. Will some of the world’s leading climate scientists be prevented from contributing to the Intergovernmental Panel on Climate Change’s Summary for Policymakers because they are dependent on government money? Should British wildlife experts not give policy advice to foreign nations attempting to save biodiversity? Surely not, but this is one possible reading of the clause.

Scientists in the United Kingdom could be forgiven for feeling baffled by the development, given that the government-funded research councils have spent recent years promoting the ‘impact agenda’. This encourages scientists to make sure that their work has reach outside their own academic disciplines, including influencing policy and legislation.

Officials have indicated that the problem can be fixed. Ministers have the power to remove the rule entirely from grants, or add in ‘qualifications’ that could permit some limited additional uses for the money.

All researchers supported by government — regardless of what organizational auspice they operate under — should be in no doubt that they have not only a right but a duty to speak out about the implications of their work. There must be a complete exemption of any research from this clause, not just for those who work in academia, but for those who work directly for government.

*Nature* **531**, 7 (03 March 2016) doi:10.1038/531007a

---

*Tweet Follow @NatureNews*

---

## Related stories and links

### From nature.com

- **Confusion reigns as UK scientists face government ‘gagging’ clause**  
26 February 2016
- **Canadian election brings hope for science**



20 October 2015

- **Communication breakdown**

01 April 2015

---

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

## 1 comment

**Subscribe to comments**



Robert Ward • 2016-03-02 03:56 PM

A petition has been launched to persuade the UK Government to exempt university researchers from the new 'anti-lobbying' clause. You can sign it here:

<https://petition.parliament.uk/petitions/122957>

---

## See other News & Comment articles from *Nature*

**Nature** ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.  
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER



# Future present

**A young global-sustainability platform deserves time to find its feet.**

01 March 2016

In pure science, as in art, little is urgent. Gravitational waves were discovered — a triumph for curiosity-driven science — thanks to physicists' patience and imaginative power. That they had waited decades is irrelevant. Alas, not all science has the luxury of timelessness.

Urgent science touches on issues that rank high on the social agenda. Theorists have classified fields such as climatology and global-change research as post-normal science, in which socio-economic stakes are high and decisions are pressing. That is the case with the agenda of Future Earth, an international sustainability-research platform set up in 2012 to tackle complex social challenges, from climate change to finance.

The scheme replaces a number of narrower programmes, including the international Geosphere-Biosphere and Human Dimensions programmes, which — to the regret of many — are now all closed.

Sustainability researchers will need to follow a multidisciplinary— nay, transdisciplinary — approach that goes beyond what many scientists have been used to. Future Earth's 'co-design' intends natural and social scientists to plan and carry out research with outside experts. Whether that will win over academic researchers, stakeholders and, crucially, funders remains to be seen. To convince sceptics, the scheme needs to provide a successful example of how it will work in practice.

Preservation of the natural commons, such as atmosphere, water, land and oceans, for future generations is vital and a cause to which any responsible scientist will happily subscribe. But combining the conventional scientific methods of the natural and social sciences with knowledge from various other sources — land owners and planners, insurance companies, conservation groups, emergency organizations and political decision-makers — poses conceptual and organizational challenges.

A cross-community Future Earth workshop on adaptation and responses to extreme climate events, held last month in Berlin, offered a taste of such challenges (see [go.nature.com/6utfmi](http://go.nature.com/6utfmi)) and might serve as a test for the design of research networks on sustainability issues. Under time pressure, participants had to draft a research strategy to address the drivers and implications of extreme events, and make it fit with Future Earth's conceptual framework — a tough issue. The workshop asked scientists from different academic cultures to do that work, which produced semantic confusion and the odd unhelpful generalization.

## Related stories

- Adaptation trade-offs
- Finite Earth
- The political economy of climate adaptation

## More related stories



But the workshop was not in vain. Many participants (a healthy number of whom were from developing nations) said that they revelled in being pushed out of their comfort zones. They produced several meaty research questions, including some genuinely new ideas for how the social and natural sciences could interact. For example, when do climatic and socio-economic factors combine to amplify the impacts of climate extremes and induce cascading harm? Are there 'tipping points' at which social or natural systems might fail to recover from shocks? And how might science-based adaptation work in data-scarce regions?

The ideas found an audience. Representatives of funding agencies at the workshop cautiously indicated that the proposals stand a good chance of getting funded by the Belmont Forum, a worldwide group of 21 major funders of global environmental-change research.

But governments and grant-giving agencies have not yet firmly committed to funding Future Earth as a whole. The reluctance comes from uncertainty over what the scheme might be able to deliver. The closure of successful programmes in favour of something fashionable but conceptually unproven has earned Future Earth sceptical glances. But then, it was launched in response to complaints that previous programmes were not sufficiently linked and that the knowledge they produced was scarcely picked up in practice. There is no lack of studies, for example, on how extreme heat, rain and wind affect farmers, city dwellers and coastlines in many parts of the world. But the results are almost useless if they never make it beyond the pages of academic journals.

Future Earth will need to make sure that scientific evidence comes to the desks of decision-makers, no matter what they might then make of it. But the programme should also avoid retreading familiar ground. The mountain of data from previous programmes, including countless climate-change studies, remains relevant — even if the information hasn't yet been put to constructive use.

Future sustainability research, no matter how interdisciplinary, should build on that heritage and focus on finding and closing knowledge gaps. In doing so, scientists involved in Future Earth can provide an invaluable service to society. And researchers in niche disciplines — palaeoclimatology or behavioural science, say — who work to fill those gaps will get a welcome chance to put their work into a broader context.

Future Earth might also become a showcase for linking natural and social sciences — a real necessity given that human activity is altering the planet at worrying speed. But sustainability research must not become tied in the straitjacket of conceptualism and utilitarianism. Scientists are not merely service providers. As in any other field of science, sustainability research must remain at its core a curiosity-driven affair.

*Nature* **531**, 7–8 (03 March 2016) doi:10.1038/531007b

---

*Tweet Follow @NatureNews*

## Related stories and links

---

### From nature.com

- **Adaptation trade-offs**  
23 October 2015
- **Finite Earth**  
21 September 2015
- **The political economy of climate adaptation**  
24 June 2015
- **Coastal conundrums**  
28 January 2015
- **Climate panel says prepare for weird weather**  
18 November 2011
- **Climate and weather: Extreme measures**  
07 September 2011

### From elsewhere

- **Future Earth**
  - **UN Sustainable Development Goals**
  - **Extreme Events and Environments – from climate to society**
  - **International Geosphere Biosphere Programme (closed)**
  - **International Human Dimensions programme (closed)**
- 

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

## Comments

**Subscribe to comments**

There are currently no comments.

### See other News & Comment articles from *Nature*

**Nature** ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.  
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER



Sustainability researchers will need to follow a multidisciplinary — nay, transdisciplinary — approach that goes beyond what many scientists have been used to. Future Earth's 'co-design' intends natural and social scientists to plan and carry out research with outside experts. Whether that will win over academic researchers, stakeholders and, crucially, funders remains to be seen. To convince sceptics, the scheme needs to provide a successful example of how it will work in practice.

Preservation of the natural commons, such as atmosphere, water, land and oceans, for future generations is vital and a cause to which any responsible scientist will happily subscribe. But combining the conventional scientific methods of the natural and social sciences with knowledge from various other sources — land owners and planners, insurance companies, conservation groups, emergency organizations and political decision-makers — poses conceptual and organizational challenges.

A cross-community Future Earth workshop on adaptation and responses to extreme climate events, held last month in Berlin, offered a taste of such challenges (see [go.nature.com/6utfmi](http://go.nature.com/6utfmi)) and might serve as a test for the design of research networks on sustainability issues. Under time pressure, participants had to draft a research strategy to address the drivers and implications of extreme events, and make it fit with Future Earth's conceptual framework — a tough issue. The workshop asked scientists from different academic cultures to do that work, which produced semantic confusion and the odd unhelpful generalization.

But the workshop was not in vain. Many participants (a healthy number of whom were from developing nations) said that they revelled in being pushed out of their comfort zones. They produced several meaty research questions, including some genuinely new ideas for how the social and natural sciences could interact. For example, when do climatic and socio-economic factors combine to amplify the impacts of climate extremes and induce cascading harm? Are there 'tipping points' at which social or natural systems might fail to recover from shocks? And how might science-based adaptation work in data-scarce regions?

The ideas found an audience. Representatives of funding agencies at the workshop cautiously indicated that the proposals stand a good

chance of getting funded by the Belmont Forum, a worldwide group of 21 major funders of global environmental-change research.

But governments and grant-giving agencies have not yet firmly committed to funding Future Earth as a whole. The reluctance comes from uncertainty over what the scheme might be able to deliver. The closure of successful programmes in favour of something fashionable but conceptually unproven has earned Future Earth sceptical glances.

**“Sustainability research must not become tied in the straitjacket of conceptualism.”**

But then, it was launched in response to complaints that previous programmes were not sufficiently linked and that the knowledge they produced was scarcely picked up in practice. There is no lack of studies, for example, on how extreme heat, rain and wind affect farmers, city dwellers and coastlines in many parts of the world. But the results are almost

useless if they never make it beyond the pages of academic journals.

Future Earth will need to make sure that scientific evidence comes to the desks of decision-makers, no matter what they might then make of it. But the programme should also avoid retreading familiar ground. The mountain of data from previous programmes, including countless climate-change studies, remains relevant — even if the information hasn't yet been put to constructive use.

Future sustainability research, no matter how interdisciplinary, should build on that heritage and focus on finding and closing knowledge gaps. In doing so, scientists involved in Future Earth can provide an invaluable service to society. And researchers in niche disciplines — palaeoclimatology or behavioural science, say — who work to fill those gaps will get a welcome chance to put their work into a broader context.

Future Earth might also become a showcase for linking natural and social sciences — a real necessity given that human activity is altering the planet at worrying speed. But sustainability research must not become tied in the straitjacket of conceptualism and utilitarianism. Scientists are not merely service providers. As in any other field of science, sustainability research must remain at its core a curiosity-driven affair. ■

## Brain power

*As brain stimulation finds non-medical uses, now is the time to consider its implications.*

It is a cautionary tale for twenty-first-century medicine. Late last year, neurosurgeons in the United States reported odd symptoms in three of their patients. Well into their sixties and seventies, these people complained of headaches, nausea, unstable balance, weak legs, low blood pressure and falling down. Chest X-rays revealed the problem. Two devices implanted into their chests — a pacemaker to help their failing hearts and a battery unit that powered electrodes buried deep inside their heads to control the signature tremors of Parkinson's disease — had been placed too close together. One machine was interfering with the functioning of the other (M. Sharma *et al.* *Basal Ganglia* 6, 19–22; 2016).

From iron lungs and dialysis machines to implantable defibrillators, we are used to technology helping our bodies. Deep-brain stimulation — the electrodes and battery implanted in the patients' heads — has been helping people with neurological and psychiatric disorders for more than a decade, but it requires quite a commitment. Brain surgery is expensive and not for everybody. The number of people who might benefit is very small given the overall burden posed by mental illness and related problems. This is one reason why there is a lot of interest in cheaper and easier types of brain stimulation, which apply electric current and magnetic fields to the outside of the head.

If these types of brain stimulation are found not to produce much of a difference, then it will not have been for a lack of effort. Academic

journals are filling up with case reports and preliminary trials of the technologies to help people with depression, autism spectrum disorders, schizophrenia, obsessive-compulsive disorder, addiction, anxiety and many more cognitive problems. It is early days, but there are enough positive results to draw the attention of people who struggle with such issues or know someone who does. Some of these people want to try it on themselves or their children. An electrical brain stimulator is fairly simple to make, and even simpler to buy from one of the companies that are popping up to sell them online. Self-medication has never been so high-tech.

Many neuroscientists have raised the alarm over do-it-yourself (DIY) brain stimulation, pointing out that it can be unsafe in the short term and might have side effects in the long term. Some want regulation. But there is another, more fundamental, ethical issue that must be confronted. As we report in an Outlook article on page S6 — one of a series of pieces that discuss cognitive enhancement — the use of DIY brain stimulators is not confined to those for whom conventional medicine has failed. A small but growing number of people want to use the devices to improve their natural mental abilities. And in so doing, this community is piggy-backing on scientific studies that suggest that electric currents and magnetic fields could improve academic performance by boosting memory and attention, and perhaps even alter attitudes.

The use of medicines to enhance performance in sport is frowned on, and a clear line has been drawn between taking them to treat and taking them to cheat. Could a similar distinction be made for cognitive-

enhancement techniques? Should it be? It's too soon to answer some of these questions — scientists and doctors must first reach consensus on the effectiveness of the techniques — but it is not too soon to ask them. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunqv](http://go.nature.com/xhunqv)



## Stop needless dispute of science in the courts

*Primers on various scientific topics could be used across trials to avoid wasting time on debating basic points, argues David Neuberger.*

Testimony from expert witnesses — and I have heard a lot in my career as a judge — is a long-standing and important feature of legal proceedings. The scientists, engineers, inventors and technologists who offer their opinions in court are encouraged to agree on basic points before a trial begins. But they often do not agree as much as we hope. That tends to lengthen the time taken to cross-examine them, and contributes to justice being an expensive, drawn-out and stressful experience for all involved.

Better would be for courts to have a set of scientifically agreed principles that lay out the consensus opinion on some topics, and where there is reasonable doubt on others. Judges typically get such a primer when trying a patent dispute. Both sides allow their expert witnesses jointly to present points on which they agree, and which will not be disputed. This effectively sets a baseline for the ensuing arguments, which can still diverge significantly. These primers are useful, but only for specific cases. When the case they were prepared for is settled, the primer normally becomes redundant.

It's not realistic for primers to be prepared individually for every case, but perhaps they could be created for topics that recur — and that are argued about each time. Scientific issues arise in a substantial number of cases, certainly enough to justify a primer that could be applied to many of them. These could be broken down into four themes. For example, forensic-science primers could detail how crime-scene samples can be matched to DNA profiles and how mixed profiles can be disentangled. Pure-science primers could explain how computer memories can be accessed and interpreted, and good-practice primers could lay out appropriate medical treatments and techniques. Scientific-method primers could set out the use and reliability of statistics.

From my experience, such primers would be hugely beneficial. They should set out the current generally accepted facts and opinion in each area, and be written as far as possible in accessible language (as all evidence in a court should be, but alas not always is).

Such primers would save money and time because the issues they detail would not realistically be open to challenge. They would also help in assessing the reliability of expert witnesses who give evidence on such issues, and they would increase the proportion of cases that are settled without a trial. The fact that opinions that are generally accepted in the scientific world sometimes turn out to be wrong is no barrier to this proposal. It is an inherent risk in giving and weighing up scientific evidence.

The legal process is not static, and courts are already working on new ways to test the evidence of expert witnesses. Many critics think

that formal cross-examination risks the court favouring a more-fluent witness or a cleverer cross-examiner rather than the best evidence. One possibility, already being adopted as an alternative to cross-examination in some civil litigation, involves so-called concurrent evidence (or 'hot-tubbing' as it is colloquially known), in which the experts and lawyers sit around a table and discuss the issues at a relatively informal, if structured, meeting that is chaired and led by the judge. The scientific primers that I have suggested build on this approach.

How could they be prepared? It would require identifying areas of expertise in which a primer would be helpful and feasible, and then getting a group of acknowledged experts to formulate the guidance in that area. It would also, I think, be necessary for the group to monitor the primer, to take into account both how it is working and what advances are being made in the area.

This would involve the legal and scientific communities working closely together, which is already starting to happen. As part of broader discussions, I and other senior judges are talking to scientists and officials at the Royal Societies of London and Edinburgh on how they could help us to prepare primers. We hope to announce some progress soon.

The law has much to learn from science, in terms of both scientific thinking and discoveries and inventions. Scientific thinking is inevitably different from legal thinking — the idea of what constitutes proof and the role of common sense are two examples of divergence. But, given the importance of experience, logic and humanity in both spheres, legal and scientific thought have much in common as well.

As for scientific advances, they interrelate with law both specifically (patents, for example) and generally (DNA evidence). And, as scientific research improves our understanding of the brain and mental processes, science will have even more to offer the law on issues such as mental capacity, the extent of pain and the reliability of memory.

It is not a one-way relationship. As scientific discoveries and inventions continue to move into ethically controversial territory, the law will be able to provide a clear and robust framework to accommodate such developments. Two examples include the relationship between surveillance and privacy, and genetic engineering.

More broadly, lawyers and scientists who learn from each other's expertise and experience can benefit society as a whole. Such a relationship of mutual cooperation is one of which I am sure that Francis Bacon, the remarkable jurist, scientist and essayist who died 390 years ago, would have wholeheartedly approved. ■

**David Neuberger** is president of the UK Supreme Court in London.  
e-mail: jackie.sears@supremecourt.uk

**PRIMERS WOULD  
SAVE MONEY  
AND TIME  
BECAUSE THE ISSUES  
THEY DETAIL WOULD  
NOT REALISTICALLY  
BE OPEN TO  
CHALLENGE.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/39ebk8](http://go.nature.com/39ebk8)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## HUMAN EVOLUTION

### Interbreeding in ancient Africa

Fossils from Africa exhibit features of both modern humans and archaic species — possible evidence for interbreeding. However, finding a genetic legacy of such encounters has been difficult because of a lack of ancient human DNA from Africa.

Michael Hammer at the University of Arizona in Tucson and his team analysed the entire genomes of seven individuals from two contemporary Western African Pygmy groups. The researchers identified 265 regions of the Pygmy genomes that may have been acquired through ancient interbreeding with other species, events that could have happened as recently as 9,000 years ago.

**Genome Res.** <http://doi.org/bct3> (2016)

## DEVELOPMENTAL BIOLOGY

### Gene lets animals tell left from right

A gene defines left and right during embryonic development in snails and frogs.

Animals generally look symmetrical, but internal organs are often positioned asymmetrically. To find out how embryos first define left and right at the molecular level, Angus Davison at the University of Nottingham, UK,



and his colleagues compared the DNA of pond snails (*Lymnaea stagnalis*; pictured) that had shells with clockwise or anti-clockwise spirals. They found that formin, a cell-structure protein, was consistently linked to spiral direction and is expressed early in snail development, showing asymmetry even in two-cell embryos.

The team treated frog embryos (*Xenopus laevis*) with anti-formin drugs, and found that 13% developed an organ on the opposite side to its normal position, suggesting that formin also coordinates

this process in frogs.

**Curr. Biol.** <http://dx.doi.org/10.1016/j.cub.2015.12.071> (2016)

## CLIMATE SCIENCE

### Where climate models fall short

Climate models tend to overestimate the extent to which climate change contributes to weather events such as extreme heat and rain.

Omar Bellprat and Francisco Doblas-Reyes at the Catalan Institute of Climate Sciences in Barcelona, Spain, used an idealized statistical model

to compare the frequency of weather extremes in simulations with and without climate warming. Extreme events seemed to be more closely linked to climate change when the model was forced to run at low levels of reliability than when the model error was kept to a minimum.

To account for models' biased representation of climate variability, studies should rely on calibrated model ensembles, which are commonly used by weather forecasters, the authors suggest.

**Geophys. Res. Lett.** <http://doi.org/bcsr> (2016)



## ECOLOGY

### Hearing is fearing for raccoons

Fear of predators can trigger a cascade of effects through an ecosystem.

Humans have eliminated most predators of raccoons (*Procyon lotor*; pictured) — such as cougars and wolves — from the Gulf Islands of British Columbia in Canada. This has allowed the raccoons to forage almost freely on shoreline species such as crabs and fish. To instil fear in the raccoons, Justin Suraci at the University of Victoria in Canada and his colleagues broadcast dog vocalizations over various island shores

for one month. They found that foraging by raccoons decreased drastically at these locations compared to areas where seal vocalizations were broadcast. This caused the number of some crabs to increase by up to 97% and numbers of some fish to rise by 81%. The snail prey of one crab species saw declines.

This manipulation of fear shows the cascading effects of losing large predators from ecosystems, the authors say.

**Nature Commun.** 7, 10698 (2016)

SHANNA BAKER/HAKAI MAGAZINE

ESTHER DE ROU/GARY MCDOWELL/  
JEREMY GUAY, PEREGRINE CREATIVE



## NEUROSCIENCE

## Light rewrites memories of place

Researchers have used light to disconnect the memory of an experience from that of the location where it occurred.

'Place' cells in the brain fire when the animal is in a particular location, helping it to remember that place. Stéphanie Trouche and David Dupret at the University of Oxford, UK, and their colleagues wired up the brains of mice to monitor these cells and switch them off using light. When the researchers switched off the place cells that were associated with one of two differently shaped enclosures, they found that a group of previously silent place cells fired instead to encode that location.

The team injected mice in one enclosure with cocaine, which made the animals prefer that location. When the initially active place cells were switched off, the mice no longer sought out the cocaine-linked enclosure — yet behaved as if it were still familiar.

*Nature Neurosci.* <http://doi.org/bcscv> (2016)

## CHEMISTRY

## Catalyst for clean drinking water

An efficient and affordable catalyst could improve access to clean drinking water in remote areas.

Hydrogen peroxide is commonly used to treat water and as a disinfectant, but it is synthesized in large-scale facilities at high concentrations that require dilution before use. Simon Freakley and Graham Hutchings at Cardiff University, UK, and their colleagues created a series of catalysts that can be used to make small batches of diluted hydrogen peroxide directly from hydrogen and oxygen.

An earlier version of their catalyst used gold and palladium supported on activated carbon. Their latest version replaces gold

with cheaper materials, including tin, zinc and nickel, but it maintains the same high reaction efficiency of more than 95%, and uses commercially available support materials such as titanium dioxide.

*Science* 351, 965–968 (2016)

## CANCER

## Immune changes drive metastasis

Quantifying the number of cancer-fighting immune cells that a tumour contains could offer a way to predict whether it will spread through the body.

Cancer is often deadly when it spreads, but anticipating primary-tumour spread (or metastasis) is difficult. Jérôme Galon at the French National Institute of Health and Medical Research in Paris and his team analysed tumours from more than 800 people with colorectal cancer, comparing people whose tumours had metastasized with those whose had not. The primary tumours from both groups had similar patterns of mutations in cancer genes, but tumours that had spread had fewer cell-killing T cells. The invasive edges of the metastasized tumours also had a lower density of lymphatic vessels, which transport immune cells.

The authors conclude that these changes contribute to metastasis, and that immunotherapies that boost T-cell responses could block the spread of cancer in people with early-stage disease.

*Science Transl. Med.* 8, 327ra26 (2016)

## COSMOLOGY

## Missing matter may hide in voids

As much as 30% of the Universe's observable matter could be hiding in enormous cosmic voids, where it is too sparse for scientists to observe.

Matter in the nearby Universe is said to be missing because astronomers have failed to see as much material

## SOCIAL SELECTION

Popular topics on social media

## How many replications are enough?

When psychologist Courtenay Norbury came across a paper in *Research in Developmental Disabilities* this week that had similar conclusions to research she published 12 years ago, she turned to social media with a question. Norbury, who studies children with autism spectrum disorders at University College London, tweeted: "How many times does a research finding need to be replicated before the field says 'ok, how do we move this forward?'" Dorothy Bishop, a developmental neuropsychologist at the University of Oxford, UK, who helped to write a report on how to improve the reliability of biomedical research, tweeted in response

that some fields can get stuck on the same research questions: "The opposite of the reproducibility crisis! Stasis. And yup it's a problem in some areas."

➔ **NATURE.COM**  
For more on popular papers:  
[go.nature.com/9yaf4d](http://go.nature.com/9yaf4d)

*Res. Dev. Disabil.* <http://doi.org/bctr> (2016)

as observations of the early Universe suggest there should be. To map how matter might be distributed, Markus Haider at the University of Innsbruck in Austria and his team used a simulation for how galaxies and intergalactic filaments evolved. This modelled the behaviour of both normal and dark matter — an invisible substance detected only by its gravitational pull — in a cube of space 350 million light years (about 107 million parsecs) across.

Analysis of the model, known as Illustris, suggests that the energy of radiating supermassive black holes has flung as much as 24% of normal matter out of galaxies and into voids, where an extra 6% that has yet to gather in filaments also lies. This could help to explain some of the missing matter, say the authors.

*Mon. Not. R. Astron. Soc.* 457, 3024–3035 (2016)

A team led by Richard Levy of GNS Science in Lower Hutt, New Zealand, analysed a drill core of sediment from McMurdo Sound, Antarctica, to reveal climate history between 21 million and 13 million years ago. The greatest ice-sheet shrinkage was seen when CO<sub>2</sub> levels were 500 parts per million or more: today's level is about 400 p.p.m., and rising. The researchers conclude that Antarctica (pictured) may respond more quickly to changing CO<sub>2</sub> levels than once thought.

A related study by some of the same authors modelled how the Antarctic ice sheet responded to shifts in climate and found similar changes. *Proc. Natl Acad. Sci. USA* <http://doi.org/bcv4> (2016); <http://doi.org/bcv5> (2016)



## CLIMATE SCIENCE

## Sample reveals Antarctic history

The Antarctic ice sheet retreated inland millions of years ago, when atmospheric carbon dioxide levels were not that much higher than they are now.

➔ **NATURE.COM**  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)

# How many replication studies are enough?

Researchers on social media ask at what point replication efforts go from useful to wasteful.

Dalmeet Singh Chawla

26 February 2016

When psychologist Courtenay Norbury came across a paper this week that had similar conclusions to research she published 12 years ago, she turned to social media with a question. Norbury, who studies children with autism spectrum disorders at University College London, tweeted:

How many times does a research finding need to be replicated before the field says "ok, how do we move this forward?"

— Courtenay Norbury (@lilacCourt) February 23, 2016

Print

Dorothy Bishop, a developmental neuropsychologist at the University of Oxford, UK, who helped to write a report on how to improve the reliability of biomedical research, tweeted in response that some fields can get stuck on the same research questions:

the opposite of the reproducibility crisis! Stasis. And yup it's a problem in some areas <https://t.co/tzXMUQ6h8C>

— Dorothy Bishop (@deevybee) February 23, 2016

The problem of irreproducibility in science has gained widespread attention, but one aspect that is discussed less often is how to find the right balance between replicating findings and moving a field forward from well-established ones. Norbury has for years reported that people who have autism and poor language skills can find it difficult to make inferences, decipher ambiguous phrases and understand metaphors or jokes. She has also shown that this is not necessarily the case for people with autism who are more linguistically capable<sup>1, 2</sup>. The latest study, published in *Research in Developmental Disabilities* on 18 February<sup>3</sup>, mirrored the work by suggesting that people with autism who have good language skills do well on such tasks.

Psychologists Melanie Eberhardt at the University of Cologne in Germany and Aparna Nadig at McGill University in Montreal, Canada, authors of the latest study, acknowledge that this evidence is clear and convincing for researchers in this area, but say: "We find that unfortunately there are many lay, professional and academic circles where this result is still not understood." They add, "Therefore we found it important to add to the convergent evidence on this question."

"Replication is important but it would be nice sometimes to take a bigger leap forward," says Norbury. She adds that the next "obvious" step in this research involves intervention, which is challenging to do. "But I'd love to start thinking of how to overcome these obstacles, rather than just repeatedly demonstrating that language impairment has negative impacts for children with autism," she says.

Brett Buttlere, a research assistant at the Leibniz Institute for Knowledge Media in Tübingen, Germany, tweeted:

@deevybee this is also a large problem in Psychology as well, in my opinion! :D



## Social Selection

Nature's snapshot of science on social media

### Related stories

- Over half of psychology studies fail reproducibility test
- Collaborate and listen to reproduce research
- Irreproducible biology research costs put at \$28 billion per year

### More related stories

— Brett (@BrettButtliere) February 23, 2016

“It is obvious that not publishing when something doesn’t work is bad, but so is doing the same thing over and over again without learning something new,” Buttliere said in an interview.

Virginia Barbour, executive officer of the Australasian Open Access Support Group in Brisbane, Australia, posted her observations:

@deevybee @lilacCourt I've seen meta analyses of clinical interventions which show that there was clinical certainty yet trials continued

— Ginny Barbour (@GinnyBarbour) February 23, 2016

In response, Bishop later tweeted a link to a paper published in *The Lancet* that looked at reducing waste in biomedical research<sup>4</sup>. The 2014 article said that many studies are done without referencing systematic reviews of the literature, which leads to waste.

Paul Glasziou, a clinician and researcher at Bond University in Queensland, Australia, who is a co-author of the *Lancet* paper, says that the bar for reproducibility can be set at different heights. For example, he says, the US Food and Drug Administration requires a minimum of two positive randomized control trials to show effectiveness of a new drug — a rule that Glasziou says is “reasonable” as long as the trials are “well done and adequately powered”. He adds, however, that clinical studies are usually repeated more than once, pointing to one analysis of systematic reviews that found that a review cited on average 16 similar papers<sup>5</sup>.

One issue could be that researchers are unaware of similar studies. For example, a study of 1,523 clinical-trial reports published between 1963 and 2004 found that, on average, each report cited less than 25% of the previous similar trials that were relevant<sup>6</sup>.

Barbour added in an interview that one way of deciding whether a claim has been reproduced enough could be to analyse reviews and meta-analyses of studies from the same field.

Using novelty as a criterion for publication in journals may solve the problem, notes Norbury, who is an editor of the *Journal of Child Psychology and Psychiatry*. By contrast, many researchers have called for journals to de-emphasize new findings and instead publish numerous replications; Norbury says that she agrees “to a certain extent”.

But, she adds, “my heart sometimes sinks when I see yet another paper exploring what I consider to be well-trodden ground. I’m not looking for ‘sexy’ findings, but I am looking for something that has the potential to change practice or move the field forward.”

*Nature* **531**, 11 (03 March 2016) doi:10.1038/531011f

References

1. Norbury, C. F. J. <i>Speech Lang. Hear. Res.</i> <b>47</b> , 1179–1193 (2004)	
Show context	Article
2. Norbury, C. F. J. <i>Exp. Child Psychol.</i> <b>90</b> , 142–171 (2005)	
Show context	Article
3. Eberhardt, M. & Nadig, A. <i>Res. Dev. Disabil.</i> <a href="http://doi.org/bctr">http://doi.org/bctr</a> (2016).	
Show context	
4. Macleod, M. R. <i>et al. Lancet</i> <b>383</b> , 101–104 (2014).	
Show context	Article PubMed
5. Moher, D., Tetzlaff, J., Tricco, A. C., Sampson, M. & Altman, D. G. <i>PLoS Med.</i> <b>4</b> , e78 (2007).	
Show context	Article PubMed
6. Robinson, K. A. & Goodman, S. N. <i>Ann. Intern. Med.</i> <b>154</b> , 50–55 (2011).	
Show context	Article PubMed



## Related stories and links

---

### From nature.com

- **Over half of psychology studies fail reproducibility test**  
27 August 2015
  - **Collaborate and listen to reproduce research**  
16 July 2015
  - **Irreproducible biology research costs put at \$28 billion per year**  
09 June 2015
  - **Reproducibility special**
- 

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

## 2 comments

[Subscribe to comments](#)

Douglas Eckberg • 2016-03-02 08:47 PM

My heart always sinks when I find people deciding it is time to micromanage others' work. The "right balance" indeed! Either the researchers of a "replication," as well as its reviewers and journal editors, are not aware of someone else's publication, in which case the importance of the original piece is probably not as great as the original writers think, or researchers, reviewers, and editors have all agreed that a fresh look is in order. In any case, this is an attempt at a "solution" without a clear "problem."

---



Leonid Schneider • 2016-02-29 12:39 PM

"Using novelty as a criterion for publication in journals", as Norbury suggests, is not such a novel idea. Most journals use this as selection criteria, with the exception of PLOS One and some others. Such "novelty"-driven approach puts the research outcome as more valuable than the research itself, and this is not at all scientific. One should not be surprised then that such fraudulent nonsense as STAP ended up in Nature: the result was very novel indeed, even if the science behind it was bonkers, and fake on top. Instead, originality of research should be assessed. This would focus on the original rationalisation, experimental setup, quality of analysis and scientific coherence (as opposed to dull and thoughtless re-pipetting of long published research). Judging the true impact of the results of these experiments should be left to peers and future scientists, not to the "goal-keeper" editors, even those elite wizards of Nature.

---

## See other News & Comment articles from *Nature*

**Nature** ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.  
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER

# SEVEN DAYS

The news in brief

## RESEARCH

### Ebola drug stutters

Results from a clinical trial to test experimental Ebola treatment ZMapp failed to show statistically significant results. The drug, developed by Mapp Biopharmaceutical, based in San Diego, California, contains three antibodies and had shown promise in animal studies. According to results presented on 23 February at the Conference on Retroviruses and Opportunistic Infections in Boston, Massachusetts, of 36 people given ZMapp, 78% survived, compared with 61% of 35 patients who did not receive the drug. Mapp was forced to end the clinical trial in January without achieving its goal of enrolling 200 patients because of the waning of the Ebola outbreak.

### Tetraquark addition

Scientists reported findings of a new tetraquark on 24 February. Elementary particles known as quarks usually bind together in groups of two or three, but physicists have observed some composed of four quarks. The new family member, called X(5568), emerged in data from the DZero experiment at the now-inactive Tevatron particle accelerator at Fermilab in

## NUMBER CRUNCH

# 947,000

The drop in Japan's population since 2010, according to the latest census. The population has fallen by 0.7%, to 127.1 million. The decline is the first since records began.



CYRIL RUOSO/MINDEN PICTURES/GETTY

## Pollinators star in biodiversity report

The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) announced the findings of its first report on 26 February. The review warns that an ongoing decline in the number of pollinating insects (**pictured**) and animals threatens global crop production, which depends on pollinators

and, as an industry, is worth up to US\$577 billion annually. According to the report, the decrease is fuelled by a multitude of factors, including climate change, disease and pesticide use. The IPBES, established in 2012, is modelled roughly on the Intergovernmental Panel on Climate Change.

Batavia, Illinois. Unlike other examples of tetraquarks, all of which contain at least two quarks of the same type, or 'flavour', each of the quarks in X(5568) is different. Studying the particle could help physicists to understand more about the strong force, which holds atomic nuclei together.

### Gas leak quantified

Some 97,100 tonnes of methane leaked out of an underground storage facility run by the Southern California Gas Company in Aliso Canyon, California, researchers reported on 25 February. A team led by Stephen Conley, president of Scientific Aviation in Boulder, Colorado, measured methane concentrations above

the site during 13 aircraft flights between 7 November and 13 February. The team calculated that the methane release was equivalent to the annual greenhouse-gas emissions from 572,000 cars. The leak began on 23 October and lasted nearly four months.

## BUSINESS

### Sequencing suit

Genome-sequencing giant Illumina said on 23 February that it has filed a lawsuit against UK-based Oxford Nanopore Technologies, the first company to commercialize nanopore sequencing. The technology reads single bases of genetic material as they pass through a nanoscale pore. The suit, by Illumina of San Diego,

California, alleges that Oxford Nanopore has infringed on Illumina patents that describe aspects of using pores to read DNA. Oxford Nanopore has its own suite of patents related to the technology. See [go.nature.com/7hydeg](http://go.nature.com/7hydeg) for more.

### Chagas scoop

KaloBios Pharmaceuticals of San Francisco, California, is poised to acquire sole distribution rights for a version of benznidazole, one of only two drugs that can treat the insect-borne parasite that causes Chagas disease, after a bankruptcy court ruled in its favour on 26 February. In December, the then chief executive Martin Shkreli announced that the company would price the drug on a level

with hepatitis C antivirals, which cost up to US\$100,000 per treatment.

## PRIZES

## Memory work wins

Three British neuroscientists share this year's Brain Prize for their work on how memories are formed and lost in the brain. Using different approaches, Timothy Bliss, visiting worker at the Francis Crick Institute in London, Richard Morris at the University of Edinburgh, UK, and Graham Collingridge at the University of Bristol, UK, have shown over the past four decades how a brain mechanism called long-term potentiation underpins the ability to learn and remember by strengthening connections between particular neurons. The €1-million (US\$1.1-million) prize was awarded on 1 March by the Grete Lundbeck European Brain Research Foundation in Denmark.

## EVENTS

## Out of deep water

A US jury has acquitted a BP site manager who was in charge of the Deepwater Horizon drilling platform during the disastrous spill in 2010, which led to 11 deaths and leaked huge amounts of oil into the Gulf of Mexico



(pictured). According to US media reports, Robert Kaluza faced a criminal charge related to the ensuing pollution, but was acquitted by a jury in New Orleans, Louisiana, on 25 February. Kaluza was one of the last BP defendants to face charges over the incident, although the company still has to pay billions in fines.

## Chemistry petition

Chemists are petitioning the chancellor of the University of California, Berkeley, to secure the future of the institution's College of Chemistry. More than 3,000 people had signed the petition as *Nature* went to press. Berkeley chancellor Nicholas Dirks announced a "strategic planning process" on 10 February, to try to find solutions to the university's "substantial and growing structural deficit". A spokesperson for the university told *Nature*

that although the College of Chemistry could be dissolved as a result of this, no decisions have yet been taken and Berkeley is committed to chemistry research and teaching.

## Italian protests

Researchers held a protest at the Sapienza University of Rome on 25 February, calling the Italian government's support for research insufficient and erratic. The protest followed a 4 February Correspondence in *Nature* by Sapienza physicist Giorgio Parisi (G. Parisi *Nature* **530**, 33; 2016) that was supported by 69 researchers. A petition to the Italian government and the European Union started by Parisi had almost 55,000 signatures as of 1 March. Italy spends 1.25% of its gross domestic product on research, but the petition says that the EU should

## COMING UP

### 7–11 MARCH

The United Nations and Costa Rica Workshop on Human Space Technology convenes in San Jose, Costa Rica.  
[go.nature.com/swmviz](http://go.nature.com/swmviz)

### 8–10 MARCH

Seattle, Washington, hosts the Climate Leadership Conference, to discuss US climate policy and innovation in the wake of the Paris agreement.  
[go.nature.com/pwplig](http://go.nature.com/pwplig)

### 10 MARCH

The US Patent and Trademark Office starts proceedings over who holds the rights to commercialize CRISPR–Cas9 gene-editing technology.  
[go.nature.com/qvnsn8](http://go.nature.com/qvnsn8)

require governments to set a minimum of 3%, as the EU Council of Ministers has advocated in the past.

## FUNDING

## India funds science

In its annual budget released on 29 February, the Indian government increased funding for the Department of Science and Technology (DST) by 17% from last year, to 44.7 billion rupees (US\$650 million). The DST is India's main funding agency and will use the money to initiate research programmes on energy, water and biomedical devices. The Department of Biotechnology received 18.2 billion rupees, a 12% rise. But news was mixed for other divisions: the Department of Health Research's budget represented a 12% rise compared to 2015–16, whereas the Department of Space got an increase of less than 2%.

➔ [NATURE.COM](http://NATURE.COM)

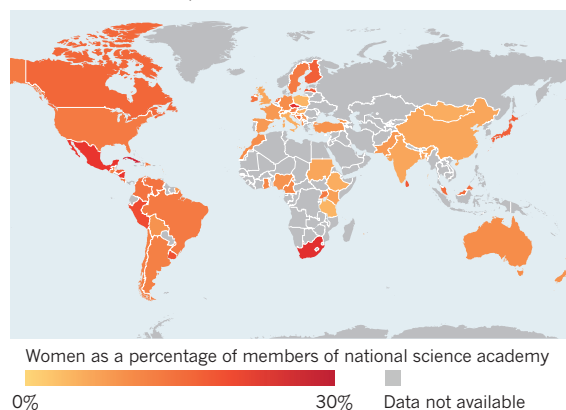
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

The first global survey of women's representation at the highest level of academia shows that just 12% of members of 69 academies surveyed in 2013–14 are female. The Cuban Academy of Sciences had the highest proportion (27%), whereas the Tanzania Academy of Sciences and the Polish Academy of Sciences had the lowest levels, at 4%. Only 40% of the academies had policies that explicitly mention the need for increased participation of women in the academy's activities. See [go.nature.com/cwigqv](http://go.nature.com/cwigqv) for more.

## WOMEN IN SCIENCE ACADEMIES

Most science academies across the world have more than 80% male membership.





# Speedier Arctic data as warm winter shrinks sea ice

Scientists push for better monitoring of what remains.

Alexandra Witze

01 March 2016

[Print](#)

UPI/Benjamin Nocerini/US Coast Guard/Eyevine

Researchers have found a way to track changes in sea-ice thickness in near real time.

Following a record winter in many ways, Arctic sea-ice cover seems poised to reach one of its smallest winter maxima ever. As of 28 February, ice covered 14.525 million square kilometres, or 938,000 square kilometres less than the 1981–2010 average. And researchers are using a new technique to capture crucial information about the thinning ice pack in near real time, to better forecast future changes.

Short-term weather patterns and long-term climate trends have conspired to create an extraordinary couple of months, even by Arctic standards. “This winter will be the topic of research for many years to come,” says Jennifer Francis, a climate scientist at Rutgers University in New Brunswick, New Jersey. “There’s such an unusual cast of characters on the stage that have never played together before.”

The characters include the El Niño weather pattern that is pumping heat and moisture across the globe, and the Arctic Oscillation, a large-scale climate pattern whose shifts in recent months have pushed warm air northward. Together, they are exacerbating the long-term decline of Arctic sea ice, which has shrunk by an average of 3% each February since satellite records began in 1979.

A persistent ridge of high-pressure air perched off the US West Coast has steered weather systems around drought-stricken California, funnelling warmth northward. As a consequence, sea ice is particularly scarce this year in the Bering Sea. “The ice would normally be extensive and cold, but we have open water instead,” says Francis.

A storm last December compounded the situation by pushing warm air — more than 20 °C above average — to the North Pole. In January, an Arctic Oscillation-driven warm spell heated the air above most of the Arctic Ocean. By February, ice had begun to circulate clockwise around the

## Related stories

- US cold snap fuels climate debate
- Summer storms bolster Arctic ice
- Special issue on the Arctic: After the ice

Arctic basin and out through the Fram Strait, says Julianne Stroeve, a researcher at the US National Snow and Ice Data Center (NSIDC) in Boulder, Colorado.

[More related stories](#)

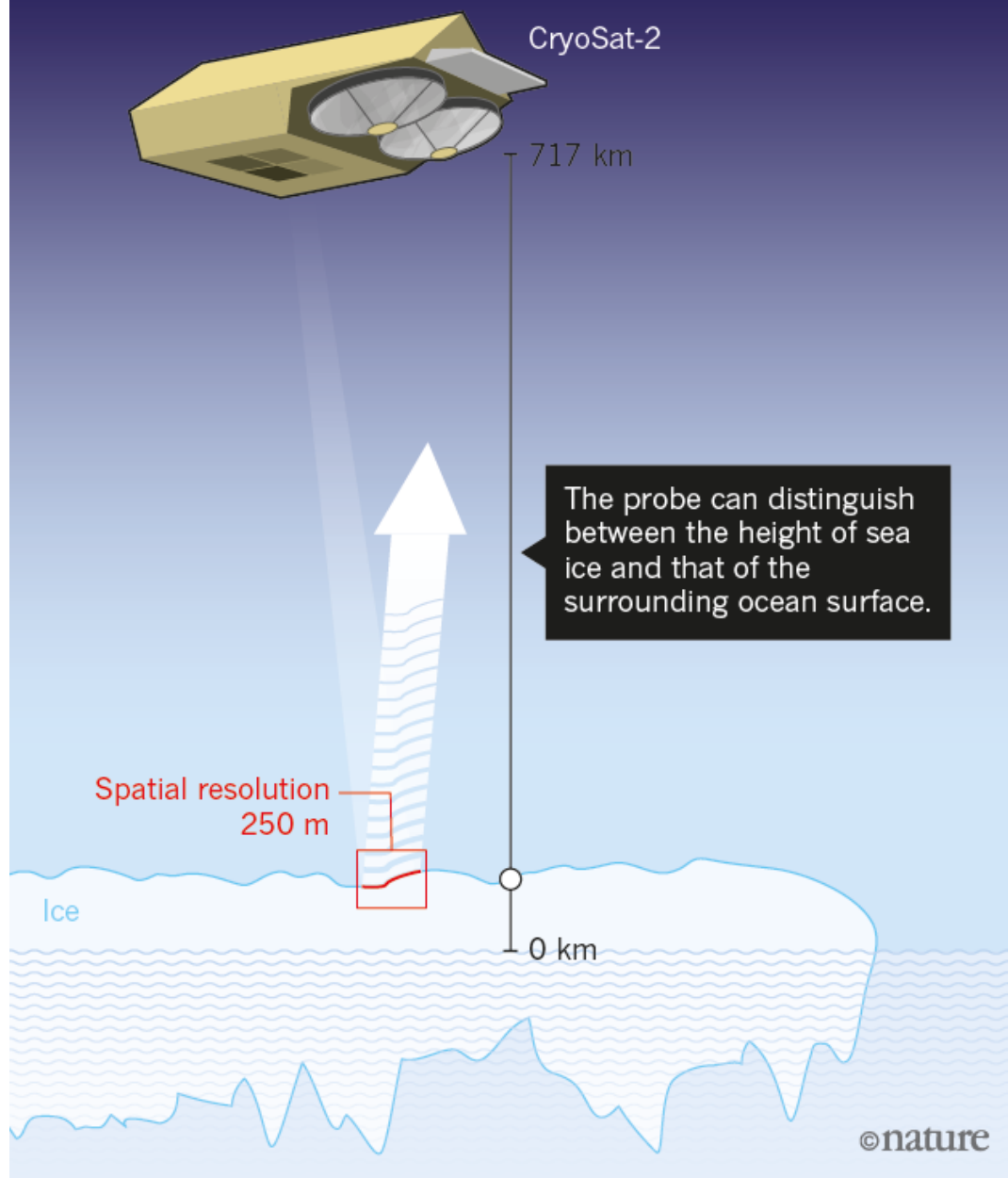
Given the Arctic's notoriously unpredictable weather, the low maximum doesn't necessarily foretell record-low melting this summer, when sea ice will reach its annual minimum. (The biggest summer melt on record happened in 2012, a year without an El Niño.) But researchers have one new tool with which to track the changes as they happen this year — the first detailed, near-real-time estimates of ice thickness, from the European Space Agency's CryoSat-2 satellite.

Three research groups currently calculate Arctic ice thickness from satellite data, but with a lag time of at least a month. Faster estimates would allow shipping companies to better plot routes through the Arctic, and scientists to improve their longer-term forecasts of ice behaviour. "The quicker you have these estimates of sea-ice thickness, the quicker you can start assimilating them into models and make more timely predictions of what's going to happen," says Rachel Tilling, a sea-ice researcher at University College London.

She and her colleagues have developed a faster way to get information on ice thickness from CryoSat-2 (see 'Measuring stick'). The satellite measures thickness by comparing the time that it takes for radar signals to bounce off the ice, as opposed to open water. Normally, it takes several months for satellite operators to calculate Cryo-Sat-2's precise orbit (and therefore the exact location of the ice and water that it flew over). But Tilling's group instead runs a quick-and-dirty analysis of orbital data, then combines it with near-real-time information on ice concentration from the NSIDC and ice type from the Norwegian Meteorological Service (R. L. Tilling *et al.* *Cryosphere Discuss.* <http://doi.org/bcw5>; 2016).

## MEASURING STICK

Radar data from the CryoSat-2 probe can now be used to track sea-ice thickness in near real time.



The result is ice-thickness measurements that are ready in just 3 days, and accurate to within 1.5% of those produced months later. The current winter cycle is the first complete season for the near-real-time data. (The measurements cannot be done in the summer, when melt ponds on the ice confuse the satellite.)

Tilling has begun to speak to shipping companies, among others, that are interested in using the data as fast as they are produced. "It really is a new era for CryoSat-2," she says.

More-accurate ice-thickness data would improve climate models and give better forecasts for the possible impacts of thick or thin sea ice, says Nathan Kurtz, a cryosphere scientist at NASA's Goddard Space Flight Center in Greenbelt, Maryland. Kurtz helps to lead NASA's IceBridge project, which will begin flying aeroplanes north of Greenland later this month to measure ice thickness using lasers and an infrared camera that can detect heat from the underlying water.

Thickness measurements are more crucial than ever, given the changing Arctic, says David Barber, a sea-ice specialist at the University of Manitoba in Winnipeg, Canada. He and his colleagues reported last year that there is increased open water all around

the edge of the Arctic ice pack every month of the year (D. G. Barber *et al. Prog. Oceanogr.* **139**, 122–150; 2015).

“We’re getting more open water in the winter than we were expecting,” Barber says. “These changes are happening very quickly, and I don’t think people are fully aware of how dramatic they are.”

*Nature* **531**, 15–16 (03 March 2016) doi:10.1038/531015a

---

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

## Related stories and links

---

### From nature.com

- **US cold snap fuels climate debate**  
08 January 2014
- **Summer storms bolster Arctic ice**  
28 August 2013
- **Special issue on the Arctic: After the ice**  
12 October 2011
- **Blog post: Arctic sea ice declines to record low**

### From elsewhere

- **Arctic Sea Ice News**
  - **Near-real-time Arctic ice thickness, from University College London**
  - **Cryosphere Today**
- 

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

## Comments

[Subscribe to comments](#)

There are currently no comments.

## See other News & Comment articles from *Nature*

*Nature* ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.  
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER



# India's budget keeps dream of genomics hub alive

Biotechnology agency wants to upgrade capabilities to kick-start economic growth.

T. V. Padma

29 February 2016 Updated: 01 March 2016



Adnan Abidi/TPX/Reuters

India's finance minister Arun Jaitley arrives at the parliament to present the latest federal budget, which contained a moderate boost for biotechnology.

An ambitious plan to turn India into a world-class centre for genomics research and commercialization received a modest boost on 29 February when the government announced its annual budget.

A big winner in the budget was India's main science funding agency, the Department of Science and Technology, which received 44.7 billion rupees (US\$650 million), a 17% hike on last year's allocation. The Department of Biotechnology (DBT), meanwhile, received 18.2 billion rupees, a 12% rise on the previous year — an indication of how the National Biotechnology Development Strategy, which the department unveiled last December, is likely to evolve.

The budget brought mixed news for other departments engaged in scientific research (see 'Budget

allocations'). The Department of Health Research's budget represented a 12% rise in funding compared with the money pledged in 2015–16, for instance, whereas the Department of Space got an increase of less than 2%.

### Budget allocations (millions of rupees)

	2015–16	2015–16 (revised)	2016–17
Department of Atomic Energy*	109,120	113,840	116,825
Defence Research & Development	143,585	124,912	135,938
Ministry of Earth Sciences	16,197	14,180	16,724
Ministry of New & Renewable Energy	3,032	2,621	50,358†
Department of Science and Technology	38,357	38,286	44,702
Department of Scientific and Industrial Research	40,310	40,367	40,628
Department of Biotechnology	16,251	16,244	18,200
Department of Space	73,880	69,594	75,091
Department of Health Research	10,181	10,126	11,448
Department of Agricultural Research and Education	63,200	55,860	66,200

\*Includes budget for operating nuclear power stations. †Includes money from a clean energy levy, now renamed the Clean Environment Cess, in contrast to previous years.

*Source: Ministry-approved budget.*

The DBT's allocation is roughly half of what Krishnaswamy VijayRaghavan, the department's secretary, estimated was needed when the strategy was released — but he is confident that the remainder can be made up from other sources. The allocation is “very good” in terms of implementing the strategy, he told *Nature*.

The DBT's strategy aims to ramp up India's total biotech revenues by more than tenfold since the industry started up two decades ago, to \$100 billion by 2025. The idea is to kick-start the economy by replicating the success of the information-technology boom that has fuelled economic growth for more than 20 years. “Biotechnology can be another vibrant model for growth that India can offer,” said the science minister, Harsh Vardhan, back in December.

India's pharmaceutical industry is largely confined to the production of ‘generic’ copies of existing drugs, and to contract research organizations, which conduct clinical trials of drugs and vaccines on behalf of pharmaceutical companies. Both have grown into successful industries, says VijayRaghavan, but the nation “is now ready to upgrade its biotech capabilities”.

The DBT received an encouraging sign last year, when its budget allocation for 2015 was not

subsequently slashed during mid-term budget revisions, as often happens. The latest 12% increase builds on that good news, says VijayRaghavan. Ahead of this year's budget, however, VijayRaghavan said that this figure would need to rise to 25–30% annually over the next five years to implement the strategy. He now says that the DBT could make up the difference with funds from elsewhere, including a national innovation mission launched in January, which identified biotechnology as a key area — in particular, to bolster the parts of the strategy aimed at nurturing biotech start-ups and supporting entrepreneurs.

The DBT could also tap into the science department's budget. Department secretary Ashutosh Sharma, who described the increase as “fantastic”, says that his department's plans include a rise in the number and quality of start-ups and business incubators, support for scientists undertaking high-risk research, the promotion of industry-relevant research at academic institutes and new research programmes on, among other things, biomedical devices.

Some of these plans overlap with the DBT's strategy. Ahead of the budget announcement, VijayRaghavan told *Nature* that the strategy revolves around three core activities. The first is the creation of new infrastructure. India already hosts several genomics research institutes, such as the Institute of Genomics and Integrative Biology (IGIB) and the National Institute of Plant Genome Research, both in New Delhi, and the National Institute of Biomedical Genomics near Kolkata. The DBT's strategy aims to create five more centres, each dedicated to a different field, including drug discovery, marine biology and infection, as well as several centres of excellence based on narrower, high-priority areas such as genetically modified organisms, vaccines and marine bioproducts.



NCBS

“India is now ready to upgrade its biotech capabilities.” Krishnaswamy VijayRaghavan, secretary of the Department of Biotechnology.

The second activity is the provision of training in the analysis of big data. Indian geneticists have previously discovered mutations in breast-cancer genes that are unique to the Indian population (M. T. Valarmathi *et al. Hum. Mutat.* **23**, 205; 2004) and sequenced the genomes of several crops, including chickpea (R. K. Varshney *et al. Nature Biotechnol.* **31**, 240–246; 2013). The DBT's strategy aims to train researchers in the scanning and analysis of large numbers of genomes. VijayRaghavan notes that the country's existing expertise in computing and biology will help this. “We need to bring the two skill sets together,” he says.

The strategy also aims to create 150 technology-transfer organizations to help to commercialize discoveries made in publicly funded research laboratories, together with 40 technology and business incubators, which will provide equipment and guidance for new firms, and facilitate networking



opportunities. Together, these all comprise the third core activity, says VijayRaghavan, and will build on the activities of the DBT's Biotechnology Industry Research Assistance Council, which was set up in 2012.

Lipi Thukral, a computational structural biologist at the IGIB, sees the DBT's goal of creating a genomics hub as an opportunity for India to enter the emerging field of precision medicine, which uses genomic, physiological and other data to tailor treatments to the individual. Achieving that will require clinicians to gather extensive data and to interact better with academics to analyse the data, says Thukral. To succeed, the strategy will also need a feedback mechanism between the biotech corporate sector and academic institutions, policies to protect scientists undertaking high-risk, high-reward research and a reorientation of academia towards more industry-relevant research, she adds.

Others have reservations about the scale of the DBT's ambitions. "The strategy seems overwhelming, overburdened, and implementation would be a Herculean task," says Nalini Vemuri, vice-president for research and development at the Gurgaon-based company Lifecare Innovations. The strategy spans four major areas — health care, food and nutrition, energy and education. But although these represent an "impressive vision", says Vemuri, a more narrowly focused goal, for example to improve the country's research in malaria and tuberculosis, would have a better chance of success.

*Nature* **531**, 16–17 (03 March 2016) doi:10.1038/531016a

---

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

## Updates

---

**Updated:** A footnote has been added to the Budget Allocations table regarding the 2016–17 figures for the Ministry of New & Renewable Energy.

## Related stories and links

---

### From nature.com

- **Indian bioscience: The anti-bureaucrat**  
13 May 2015
- **India: The fight to become a science superpower**  
13 May 2015
- **Case study: India's billion dollar biotech**  
01 August 2010
- **Biotech boom**  
27 July 2005
- **Then and now**  
27 July 2005



## From elsewhere

- **Department of Biotechnology**
  - **Biotechnology Industry Research Assistance Council**
  - **CSIR Institute of Genomics and Integrative Biology**
- 

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

## Comments

**Subscribe to comments**

There are currently no comments.

## See other News & Comment articles from *Nature*

**Nature** ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.  
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER



Ukraine has lost access to the Crimean Astrophysical Observatory.

POLITICS

# Conflicting laws threaten Ukrainian science

*Country's austerity budget stands in way of law to modernize Soviet-era academy.*

BY QUIRIN SCHIERMEIER

As political turmoil and conflict rock Ukraine, the country's main scientific organization is in a bind. In January, Parliament passed a law to modernize the ailing National Academy of Sciences of Ukraine (NASU). Yet an austerity budget imposed around the same time makes this impossible to achieve — at least this year. The resulting cuts to science funding threaten the jobs of young researchers in particular, who are best poised to revitalize the country's failing economy.

"We have an extraordinarily high number of potential young scientists who are ready to work for the welfare of the country," says Liliya Hrynevych, who chairs the Ukrainian Parliament's Committee on Science and Education and voted in favour of the modernizing law. "But without setting priorities for science and research, it will be impossible for Ukraine to become a strong and wealthy European nation."

The academy employs some 20,000 scientists across 120 research institutes. On 26 November, Parliament began to debate a "law of Ukraine on scientific and technical activity", in an attempt to streamline and strengthen the organization, which was founded in the Soviet era. Long

deemed outdated and resistant to modernization, the academy uses an opaque system to award funding, and many of its members are elderly, not least the 97-year-old metallurgist Boris Paton, who has run the NASU for decades.

The law stipulates the creation of a science advisory council that includes foreign specialists, and an independent grant-giving agency. All NASU institutes will undergo an external evaluation to examine their productivity and efficiency, and overall, government science

spending must increase from a current 0.3% of gross domestic product to at least 1.7% by fiscal year 2017 — near the European Union average.

But before the law took effect, Ukraine passed its 2016 austerity budget, in the wake of widespread closure of mines and factories, inflation, debt and currency devaluation. The budget allocates a meagre 2.05 billion hryvnia (US\$76 million) to the NASU — about 12% less than in 2015, continuing a trend of decline (see 'Ailing academy').

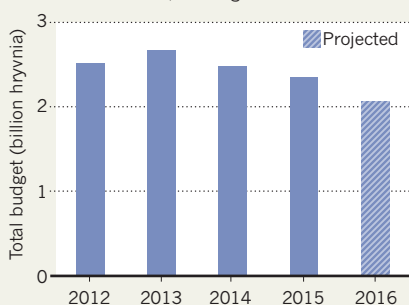
The cutbacks are irreconcilable with the science law, says Hrynevych, who is campaigning in Parliament for a budget revision after the first quarter of 2016. The budget will leave the academy with scarcely enough to cover the scant salaries (about US\$200 per month on average) paid to its administrative staff and scientists.

"We won't be able to buy any new equipment this year, and purchase of consumables will need to be reduced to a minimum," says Anatoly Zagorodny, director of the Bogolyubov Institute for Theoretical Physics in Kiev and a vice-president of the academy.

The fresh cuts, he says, will also force institutes to reduce staff — in some circumstances, by more than one-third — and to discontinue many areas of research, even though

## AILING ACADEMY

Despite the introduction of a law aimed at modernizing the National Academy of Sciences of Ukraine, funding is in decline.



SOURCE: NATL. ACAD. SCI. UKRAINE

science is crucial to economic recovery, he adds.

Young scientists are the least protected by existing labour laws and so will feel the brunt of the job cuts, says Irina Yehorchenko, a research fellow at the NASU's Institute of Mathematics in Kiev. She and some of her colleagues launched a petition in December calling on the country's president, Petro Poroshenko, to save Ukrainian science.

#### YOUTHFUL POTENTIAL

"I, for one, might be able to find a postdoc position abroad," says Oleksandr Skorokhod, a cell biologist at the NASU Institute of Molecular Biology and Genetics in Kiev who is chair of the academy's Council of Young Scientists. "But I'd much rather stay and try to change the bad state of affairs in my country."

Ukrainian science has struggled to recover from Russia's annexation of the Crimea peninsula in 2014. General consensus in the international community is that Crimea is still part of the Ukraine — the United Nations General Assembly declared invalid a March 2014 referendum in which voters in Crimea approved the peninsula's secession from Ukraine.

But all 22 Crimean institutes formerly run by the NASU are now under Russian control, and only a few of their 1,320 staff members have relocated to Ukraine-controlled territory. The academy lost access to its only research ship, the *RV Professor Vodianytsky*, three astronomical observatories in Nauchny, Katsiveli and Yevpatoria and the 204-year-old Nikitsky Botanical Garden near Yalta, on the Black Sea shore.

The Ukrainian government, moreover,

expects scientists in Ukraine to cut all ties with colleagues who stayed on the peninsula, says Hrynevych, because any collaboration would be viewed as legitimizing the Russian occupation.

The armed conflict with pro-Russian militants in eastern Ukraine is also causing problems for scientists, especially in the country's Donbas region. Some 12,000 scientists and university lecturers there — about 60% of the former staff of 26 research institutes and universities in the province — have moved to safe institutions in Kiev and elsewhere. But many evacuating scientists left behind equipment or lost irreplaceable research material. Marine, environmental and climate studies in the Black Sea region, mining-related geology and a variety of archaeological and historical research have all been hit hard, says Zagorodny. ■

#### PUBLIC HEALTH

# Spectre of Ebola haunts Zika response

*Agencies rush to show that outbreak tactics have improved.*

BY ERIKA CHECK HAYDEN

Public-health workers are still struggling to stamp out the Ebola epidemic in West Africa. But the lessons learned from that outbreak — which exposed major flaws in the global public-health system — are shaping the escalating international response to the spread of Zika virus in the Americas.

"Ebola is the gorilla in the room," says Lawrence Gostin, a health-law and policy specialist at Georgetown University in Washington DC. "It's driving everything."

He and others say that governments and international public-health agencies seem determined not to repeat the main mistake that they made with Ebola: waiting for much too long to respond to a brewing outbreak. The delay allowed Ebola to grow so out of control in West Africa that the epidemic there persists after more than 2 years and 11,000 deaths.

By contrast, the global health community has moved aggressively against Zika, beginning with a declaration from the World Health Organization (WHO) on 1 February that the clusters of microcephaly and other neurological disorders that have appeared in Brazil coincident with outbreak of the virus, and previously in French Polynesia, constitute an international public-health emergency.

The WHO has never yet made such a declaration before knowing the cause of the condition of concern. The August 2014 declaration

that Ebola was a public-health emergency came after the disease had been spreading in West Africa for 8 months and had killed 932 people. But although Zika has probably infected as many as 1 million people in the latest outbreak, the vast majority have recovered. And scientists have not proved a link between Zika and microcephaly, a condition in which infants are born with abnormally small heads and brains.

"The WHO has perhaps gotten out ahead of its usual position of gathering and verifying all the evidence before taking a clear position," says Adam Kamradt-Scott, a health-security specialist at the University of Sydney in Australia. "The WHO couldn't afford to be seen to be asleep at the wheel a second time."

Other authorities have taken similarly bold action. On 3 February, the US Centers for Disease Control and Prevention (CDC) moved its emergency-response operations centre to its highest activation level, jump-starting US government research into, and surveillance of, the Zika virus. On the same day, the United Kingdom announced the creation of a Zika research fund with an initial budget of up to £1 million (US\$1.4-million). And on 8 February, US President Barack Obama requested \$1.8 billion from lawmakers for Zika-response activities. (By comparison, Obama's \$6.18-billion request for Ebola-response funding came 3 months after that virus was declared a global emergency.)

The ongoing mobilization against Zika is not an over-reaction, says Suerie Moon,

a global-health researcher at the Harvard T. H. Chan School of Public Health in Boston, Massachusetts. Although Zika — unlike Ebola — is not usually fatal, it has the potential to cause suffering and social and economic havoc. "It's encouraging to see leadership and mobilization from WHO, CDC and other public-health institutions," Moon says. "It shows that some of the lessons from Ebola have been digested."

WHO director-general Margaret Chan has acknowledged the agency's failings on Ebola, citing "inadequacies and shortcomings in this organization's administrative, managerial and technical infrastructures" in a speech last year.

The Zika response also highlights persistent flaws of the global public-health system. Zika was first discovered in Africa in 1947, and caused a major outbreak in 2013 in the Pacific islands, but there is still no vaccine, treatment or common diagnostic test for the virus.

Kamradt-Scott wonders if the world would be tracking Zika's spread so closely if the virus had not emerged in Brazil, where hundreds of thousands of tourists are scheduled to attend the Olympic Games in August. "My own perception is that the international community hasn't responded particularly swiftly to Zika," he says.

Moon notes that although the WHO is trying to ensure that researchers in government, academia and industry share data on the outbreak, drug companies developing Zika vaccines have not publicly agreed to participate.

The WHO has long struggled to modulate its response to global-health crises, Gostin says. After it was criticized for reacting too strongly to the 2009 H1N1 influenza epidemic — declaring a full-scale pandemic, when the virus itself did not prove as deadly as was initially feared — it dialed back its response to the Ebola outbreak. Now the WHO is mounting an urgent response to Zika, in light of criticism of its reaction to Ebola. To Gostin, this inconsistency reinforces a perception that the WHO acts mainly on the basis of political, not medical, factors. "We need to stop fighting the last war," he says. ■

# Successful test drive for space-based gravitational-wave detector

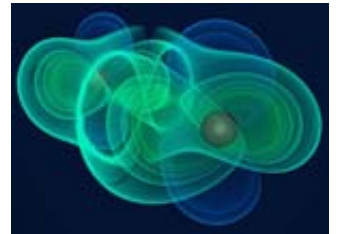
**Mission paves the way for planned €1-billion space observatory.**

**Elizabeth Gibney**

25 February 2016

Scientists have long dreamed of launching a constellation of detectors into space to observe gravitational waves — the ripples in space-time predicted by Albert Einstein and observed for the first time earlier this month.

That dream is now a step closer to reality. Researchers working on a €400-million (US\$440-million) mission to try out the necessary technology in space for the first time — involving firing lasers between metal cubes in free fall — have told *Nature* that the initial test drive is performing just as well as they had hoped.



**Nature Special:  
Gravitational Waves**

“I think we can now say that the principle has worked,” says Paul McNamara, project scientist for the LISA Pathfinder mission, which launched last December. “We believe that we now are in a good shape to look to the future and look to the next generation.”

“Everything works as we designed it. It’s sort of magical, and you rarely see that in your career as an experimentalist,” says Stefano Vitale, a physicist at the University of Trento in Italy, and a principal investigator for the Pathfinder mission.

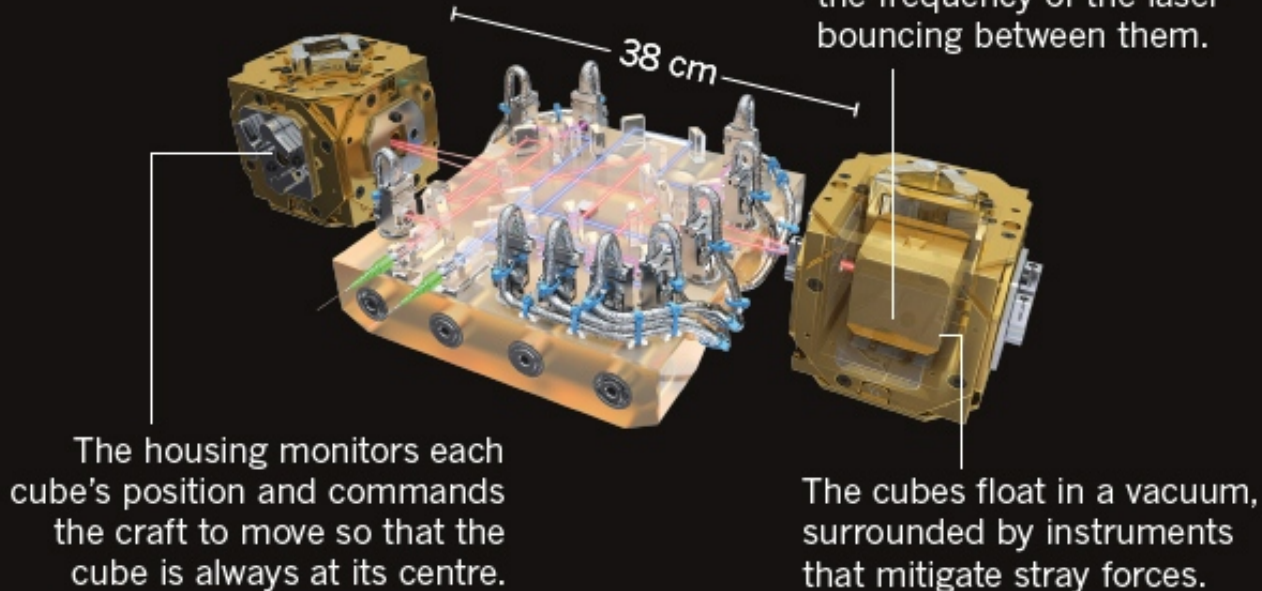


## PRECISION LAB IN SPACE

LISA Pathfinder has shown that an intricate experiment consisting of two metal cubes in freefall, isolated from all forces except gravity, can operate in space.

At the heart of Pathfinder are two free-falling metal cubes, shielded from all forces except gravity by their housing.

Any disturbance to the relative motion of the cubes affects the frequency of the laser bouncing between them.



©nature

Source: ESA/ATG medialab

The European Space Agency (ESA) financed the test, and hopes ultimately to launch a €1-billion observatory to hunt for gravitational waves. For that mission, lasers would be bounced between three spacecraft set millions of kilometres apart. Each craft would contain a test mass (a metal cube) that would be placed in free fall, protected from any forces except that of gravity. Because gravitational waves stretch and compress space-time, the observatory hopes to be able to see passing waves by using the lasers to detect minute changes in the distance between the free-falling cubes.

Because of its enormous scale, a space-based observatory could detect lower-frequency gravitational waves than can Earth-based experiments — such as the US Advanced Laser Interferometer Gravitational-Wave Observatory, which announced a first successful detection on 11 February. Lower-frequency waves can be triggered by more-powerful events, which scientists hope to study, such as collisions between galaxies and supermassive black holes.

The Pathfinder mission aimed to show — on a much smaller scale — that the basic design works, and to chart its limitations. It uses two test masses (each a 2-kilogram cube of gold and platinum) set 38 centimetres apart, floats them in isolation from everything except the influence of gravity, and tests to see whether changes in their relative movement can be measured with an accuracy of a picometre, 100,000th of the width of a human hair. To keep the cubes in free fall, the spacecraft monitors their motions and uses tiny thrusters to keep itself centred on the masses.

The complexity of such an experiment, carried out millions of kilometres from Earth, meant that sceptics doubted whether it could ever work in space, says Vitale. But data that have been streamed back since 23 February, when Pathfinder began to use the lasers to track its released cubes, show that it not only fulfils its requirements but exceeds them, he says. For now, the team is keeping under wraps details on exactly how well the instruments are performing.

Proving that the basic technology works is only the mission's first step. Its main science goal, which the Pathfinder team will work on over the coming months, is to understand where 'noise' in the system is coming from. That knowledge will be essential in designing the space-based observatory, which is scheduled for launch in 2034. "The main goal of the mission is not so much to *measure* how well we're doing, but to *understand* how well we're doing," says McNamara.

Success of Pathfinder was seen as a prerequisite for building the observatory, which ESA agreed to fund in 2013. But before launching such an ambitious experiment, scientists also considered it desirable that gravitational waves should already have been seen on Earth-based detectors. "It looks like these two conditions have been fulfilled in the same month. So it's really our month," adds Vitale.

*Nature* **531**, 30 (02 March 2016) doi:10.1038/nature.2016.19452

---

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

---

## Related stories and links

---

### From nature.com

- **Gravitational waves: How LIGO forged the path to victory**  
16 February 2016
- **Einstein's gravitational waves found at last**  
11 February 2016
- **Gravitational waves: 6 cosmic questions they can tackle**  
09 February 2016
- **Freefall space cubes are test for gravitational wave spotter**  
17 November 2015

### From elsewhere

- **LISA Pathfinder**
- 

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

## 6 comments

Subscribe to comments



antonio carlos pocob motta • 2016-03-02 02:00 PM

I believe that the gravitational Waves Must permit that can to know deeply beyond the evento horizons where the GTR might "break",new distortions in the spacetime métrics altering the speed of light,how much major the curvatures of spacetime major Must to be the deformations suffered by the speed of light?Due that breaks in GTR the singularity cannot occur,and some minimum lenght as the possible tiny of the strings might appear as fundamental vibrations and theirs harmonics admits the speed Must be variable,and the tiny of strings are spacetime,generated by the switches of left handed and right handed that are asymmetrics((particles-antiparticles)then eachtimes "particles" vibrate in a frequency only one(it is,the spacetime continuum,that is generated by particles that run forward in time and at the time run backward in time as antiparticles).Then the spacetime are tiny of strings with is harmonics,that are quantum entanglement.the in those regions of strongest energy of the black holes deform the curves of spacetime through ripples of gravitational Waves,as well as warping the quantum vaccum generating stronger fluctuations in the spacetime geometry being the topology not totally smoothes ,that are knots in the fábricas of spacetime,or foram spacetime



Pentcho Valev • 2016-02-25 07:47 PM

Variable Speed of Light, No Gravitational Waves

<http://www.damtp.cam.ac.uk/user/tong/concepts/gr.pdf> David Tong: "The rocket has height  $h$ . It starts from rest, and moves with constant acceleration  $g$ . Light emitted from the top of the rocket is received below. By this time, the rocket is travelling at speed  $v=gt=gh/c$ . This gives rise to the Doppler effect. (Neglecting relativistic effects).  $f' = f(1+v/c) = f(1+gh/c^2)$ , where  $f'$  is received frequency and  $f$  is emitted frequency." Since  $f = c/\lambda$  ( $\lambda$  is the wavelength), we have  $f' = f(1+v/c) = (c+v)/\lambda$  where  $c' = c+v$  is the speed of the light relative to the receiver (the bottom of the rocket). Clearly the speed of light (relative to the receiver) varies with the speed of the receiver, in violation of Einstein's relativity. This means that David Tong's subsequent derivation of gravitational time dilation is invalid. There is no gravitational time dilation, and accordingly there are no gravitational waves. Pentcho Valev



Pentcho Valev • 2016-02-26 08:27 AM

Clever Einsteinians know that there is no gravitational time dilation:

<http://www.printsasiasia.com/book/relativity-and-its-roots-banesh-hoffmann-0486406768>

Banesh Hoffmann: "In an accelerated sky laboratory, and therefore also in the corresponding earth laboratory, the frequency of arrival of light pulses is lower than the ticking rate of the upper clocks even though all the clocks go at the same rate. (...) As a result the experimenter at the ceiling of the sky laboratory will see with his own eyes that the floor clock is going at a slower rate than the ceiling clock - even though, as I have stressed,

both are going at the same rate. (...) The gravitational red shift does not arise from changes in the intrinsic rates of clocks. It arises from what befalls light signals as they traverse space and time in the presence of gravitation." What befalls light signals as they traverse space and time in the presence of gravitation? They accelerate of course, just as ordinary falling objects do, and this variation of the speed of light predicted by Newton's emission theory of light causes the gravitational redshift (or blueshift):

<http://courses.physics.illinois.edu/phys419/sp2013/Lectures/l13.pdf> University of Illinois at Urbana-Champaign: "Consider a falling object. ITS SPEED INCREASES AS IT IS FALLING. Hence, if we were to associate a frequency with that object the frequency should increase accordingly as it falls to earth. Because of the equivalence between gravitational and inertial mass, WE SHOULD OBSERVE THE SAME EFFECT FOR LIGHT. So lets shine a light beam from the top of a very tall building. If we can measure the frequency shift as the light beam descends the building, we should be able to discern how gravity affects a falling light beam. This was done by Pound and Rebka in 1960. They shone a light from the top of the Jefferson tower at Harvard and measured the frequency shift. The frequency shift was tiny but in agreement with the theoretical prediction." [http://www.einstein-online.info/spotlights/redshift\\_white\\_dwarfs](http://www.einstein-online.info/spotlights/redshift_white_dwarfs) Albert Einstein Institute: "One of the three classical tests for general relativity is the gravitational redshift of light or other forms of electromagnetic radiation. However, in contrast to the other two tests - the gravitational deflection of light and the relativistic perihelion shift -, you do not need general relativity to derive the correct prediction for the gravitational redshift. A combination of Newtonian gravity, a particle theory of light, and the weak equivalence principle (gravitating mass equals inertial mass) suffices. (...) The gravitational redshift was first measured on earth in 1960-65 by Pound, Rebka, and Snider at Harvard University..." And since there is no gravitational time dilation, there are no gravitational waves either. Pentcho Valev



Alone: bad. Friend: good! • 2016-02-26 09:30 AM

Do you think space has a fabric or a medium or anything other than being literally empty and nothing?



Pentcho Valev • 2016-02-25 12:41 PM

"gravitational waves – ripples in space-time first predicted by Albert Einstein" There was no such prediction. LIGO folks use the same myth:

<https://science.house.gov/sites/republicans.science.house.gov/files/documents/HHRG-114-SY-WState-DReitze-20160224.pdf> Testimony of Dr. David Reitze, Executive Director LIGO Laboratory, California Institute of Technology, Before the U.S. House of Representatives, Committee on Science, Space and Technology, on Unlocking the Secrets of the Universe: Gravitational Waves, February 24, 2016: "Let me now turn to the science of LIGO. This is what excites us the most! General relativity tells us that space-time is warped, that gravity is



geometric, and that black holes exist. These are complex concepts, deriving from a mathematically intricate but elegant theory. It also predicts the existence of gravitational waves. (...) It is worth pointing out that even though Einstein derived the existence of gravitational waves as a natural consequence of general relativity in 1916, he himself doubted they would ever be detected because the waves are so incredibly small." David Reitze misled the Committee: Einstein did not doubt "they would ever be detected because the waves are so incredibly small". Rather, Einstein doubted gravitational waves existed at all: <http://arxiv.org/abs/1602.04674>

"Around 1936, Einstein wrote to his close friend Max Born telling him that, together with Nathan Rosen, he had arrived at the interesting result that gravitational waves did not exist, though they had been assumed a certainty to the first approximation. He finally had found a mistake in his 1936 paper with Rosen and believed that gravitational waves do exist. However, in 1938, Einstein again obtained the result that there could be no gravitational waves!"

<https://www.quantamagazine.org/20160218-gravitational-waves-kennefick-interview/> ""There are no gravitational waves ... " ... "Plane gravitational waves, traveling along the positive X-axis, can therefore be found ... " ... " ... gravitational waves do not exist ... " ... "Do gravitational waves exist?" ... "It turns out that rigorous solutions exist ... " These are the words of Albert Einstein. For 20 years he equivocated about gravitational waves, unsure whether these undulations in the fabric of space and time were predicted or ruled out by his revolutionary 1915 theory of general relativity. For all the theory's conceptual elegance -- it revealed gravity to be the effect of curves in "space-time" -- its mathematics was enormously complex." Pentcho Valev



brian banton • 2016-02-25 04:51 PM

We can detect light waves travelling through the cosmic background radiation (CMB) because we possess a very efficient organ which has evolved to do so. We have no such organ to detect gravitational waves unfortunately, but that is not a sensible reason to deny their existence. However, "Ripples in the fabric of spacetime" is utter nonsense. Space is space and time is time and they do not mix, no matter how much the mathematicians would like them to. Einstein was wrong about relativity because he was not aware of the CMB and its ability to provide a universal reference frame for all light. Had he been so aware, he would have been much happier, relativity would not have existed and progress would have been much faster.

**See other News & Comment articles from *Nature***

**Nature** ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.  
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER



# Epic El Niño yields massive data trove

Waning warming event studied in unprecedented detail.

Jeff Tollefson

02 March 2016



*Reinhard Dirscherl/Ullstein Bild via Getty*

Rising ocean temperatures caused by El Niño have damaged coral in the Indian Ocean.

Floods have ravaged parts of South America. Crops are drying up in Africa. Corals are bleaching around the world. The epic El Niño warming event in the tropical Pacific Ocean has boosted temperatures and affected people and ecosystems around the globe. And thanks to a combination of luck and determination, scientists have been better placed than ever to record its evolution.

“When you have an event this big, you really want to squeeze as much out of it as you can,” says Michael McPhaden, an oceanographer at the US National Oceanic and Atmospheric Administration (NOAA) in Seattle, Washington, who helps to manage an array of buoys that is used to monitor El Niño conditions. “And we were well positioned at the beginning of 2015 to watch this thing unfold.”

Had the El Niño occurred a year earlier — as originally predicted — McPhaden’s team would not have been ready. In early 2014, the Tropical Atmosphere Ocean (TAO) array was on the verge of collapse.

NOAA eventually restored the array after oceanographers raised concerns that they — as well as weather forecasters — would be deprived of crucial data if a major El Niño arrived.

El Niño and its counterpart, La Niña — which is defined by a cooling of the equatorial Pacific — have powerful knock-on effects around the globe. This oscillation forms the basis of most seasonal weather predictions, and scientists are mining the data for clues that will improve those predictions.

NOAA has spent roughly US\$3 million to deploy aircraft, a research ship and hundreds of weather balloons to capture as much data as possible before the El Niño fades in the next few months. And the US National Science Foundation (NSF) has awarded 19 ‘rapid response’ research grants, totalling \$2.3 million, to researchers studying the event.

Many of the NSF projects focus on the effects of warm seawater on coral reefs; if water is too hot, corals bleach, expelling the colourful algae that feed them. The current bleaching event began in Guam in 2014 and has since spread to the Atlantic and Indian oceans as El Niño has warmed the seas (see ‘Selected El Niño impacts’). More than 60% of the world’s corals could be affected in the next few months, and with NOAA predicting that bleaching will continue into 2017, that number could rise.

Selected El Niño impacts

The strong El Niño warming event that began last year has affected conditions around the globe.

Eastern Africa	The failure of spring rains in 2015 affected crops in central Ethiopia and Sudan. Areas further south have been hit by heavy rains and flooding.
South America	Paraguay was hit by torrential rains this December and January, causing flooding and the evacuation of more than 100,000 people.
Australia	El Niño has been linked to warm and dry conditions in 2015.
India	Much of India experienced below-normal rainfall from June to September. There was above-normal rainfall in southern India and Sri Lanka this winter.
Oceania	Coral reefs in the central Pacific Ocean are on the front line of a global bleaching event that has also affected reefs in the Atlantic and Indian oceans.

Source: WMO

Scientists and policymakers often focus on the long-term risk posed by carbon dioxide emissions and ocean acidification. The current bleaching event — already the longest on record — has shone a spotlight on the immediate danger posed by rising ocean temperatures.

“Ocean acidification may be much less of a problem than we feared, but that’s only because many of the



corals will be dead before we get there,” says Mark Eakin, who heads NOAA’s Coral Reef Watch.

In other cases, scientists were lucky to be able to watch El Niño unfold. Daniel Rudnick, an oceanographer at the University of California, San Diego, and his colleagues received grants from the NSF and NOAA in 2012 and 2013 to deploy a trio of underwater gliders in the tropical Pacific from 2013–16. The team also released 41 new Argo floats — roughly double the previous number — along the Equator to collect temperature and salinity data down to 2,000 metres.

**“When you have an event this big, you really want to squeeze as much out of it as you can.”**

In November 2013, the gliders began to collect high-resolution measurements of the subsurface ocean current’s eastward flow. The monitoring will continue this year.

“It’s been shown that the undercurrent strengthens during El Niño, but what we have, I think, is a more finely resolved picture of its size,” Rudnick says. “It was serendipity with a capital ‘S’.”

Rudnick presented preliminary data on 23 February at an American Geophysical Union conference in New Orleans, Louisiana. The observations document the under-current’s evolution in 2014, when the forecasted El Niño fizzled, and in 2015, when it roared back. By comparing the oceanic and atmospheric conditions in both years, scientists hope to gain insights that will improve future weather forecasts.

The focus is often on the trade winds, which usually blow to the west across the Equator. Alexey Fedorov, an oceanographer at Yale University in New Haven, Connecticut, says that 2014 and 2015 both saw significant bursts of eastward winds in June, which began to push warm surface waters towards South America. In 2014, a burst of westward winds in August, driven by atmospheric patterns in the Southern Hemisphere, cut that process short — but helped to set the stage for the massive El Niño in 2015.

Fedorov says that there is no evidence that those winds could have been forecast, even by the most advanced climate models. “Sometimes we are right, sometimes we are wrong,” he says. “There was no chance to predict this.”

Although temperatures in the tropical Pacific are near their peak, McPhaden says that El Niño’s energy is quickly dissipating below the surface. “It’s very clear that this El Niño is losing its steam,” he says. And that poses another crucial question for oceanographers and climate scientists: whether this El Niño will transition into a major La Niña, as happened after the last big El Niño in 1997–98.

*Nature* **531**, 20–21 (03 March 2016) doi:10.1038/531020a

[Tweet](#) [Facebook](#) [LinkedIn](#) [Weibo](#)

## Related stories and links

---

### From nature.com

- **Monster El Niño probed by meteorologists**  
20 January 2016
- **Hunting the Godzilla El Niño**  
20 October 2015
- **Corals worldwide hit by bleaching**  
08 October 2015
- **Climate change: The case of the missing heat**  
15 January 2014

### From elsewhere

- **NOAA El Niño Campaign**
  - **Coral Reef Watch**
- 

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

## Comments

[Subscribe to comments](#)

There are currently no comments.

## See other News & Comment articles from *Nature*

**Nature** ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.  
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER



# CAN FRACKING POWER EUROPE?

*Several countries hope to unleash vast natural-gas reserves through fracking, but drilling attempts have been disappointing.*

**L**arge petroleum pumps nodded up and down in the background as British Prime Minister David Cameron donned a blue industrial jumpsuit to promote a controversial drilling technique known as hydraulic fracturing, or fracking. In his 2014 visit to a potential drill site in eastern England, Cameron laid out the benefits of tapping Britain's shale formations to release valuable natural gas. "We're going all out for shale," he said. "It will mean more jobs and opportunities for people, and economic security for our country."

Cameron hopes to replicate the surge in natural-gas production that has happened in the United States thanks to fracking — which involves injecting fluids into shale to liberate locked-up hydrocarbon deposits. The fracking revolution helped to revitalize the US economy, and Cameron's Conservative Party seeks to spark a similar gas boom in the United Kingdom. In August last year, his newly

BY MASON INMAN

elected government offered drilling licences for shale deposits and it touted estimates that "investment in shale could reach £33 billion [US\$46 billion] and support 64,000 jobs".

Over the past few years, fracking fever has swept through several European nations, including Denmark, Lithuania, Romania and especially Poland, which has seen more shale exploration than any other nation on the continent. Fracking might help to boost gas production in Europe at a time when it is facing a sharp decline. Older gas fields in the North Sea are running out, as are deposits in Germany, Italy and Romania. The disappointing output has increased Europe's dependence on imported gas, mainly from Russia. European leaders have grown wary of relying on that source, especially after diplomatic relations chilled when Russia invaded Ukraine in 2014.

But Europe's appetite for gas could increase as it tries to cut greenhouse-gas emissions — which will probably require reducing coal consumption (see 'Looming gas crunch?'). The European Commission says that "gas will be critical for the transformation of the energy system".

This means that countries such as the United Kingdom have invested an immense amount of hope in shale gas. But a close examination of the industry suggests that any fracking boom in Europe is a long way off — and some experts say that it may never arrive.

Despite several years of exploratory drilling, there are currently no commercial shale-gas wells in Europe. Tests of the region's shale potential have been limited, and the results so far have been generally disappointing, say geologists and energy experts. It remains highly uncertain how much gas would be recoverable with today's technologies, and even more difficult to forecast how much would be profitable to extract.

BARTEK SADOWSKI/BLOOMBERG/GETTY



## Fracking attempts in Poland have not led to commercial wells.

All that leads to big questions about Europe's shale hopes, says Jonathan Stern, a natural-gas expert at the Oxford Institute for Energy Studies in Oxford, UK. "There has been an enormous amount of ridiculous hype about shale gas in Europe."

## WAITING FOR A REVOLUTION

A decade ago, the United States was facing a similarly dismal outlook for natural gas. Production from conventional fields was petering out, and geologists did not expect that alternative sources of gas could compensate for the shortfall. But within a few years, the picture suddenly brightened owing to improved drilling and fracking technologies, which tapped previously inaccessible gas reserves and unleashed a boom dubbed the shale revolution. Shale is almost impermeable to oil and gas, so companies must fracture the rock to liberate those hydrocarbons.

The idea that a similar wealth of untapped energy could be lurking in the rocks below Europe is economically appealing. But geologists know relatively little about the potential of shale-rock formations in Europe because there has been less onshore drilling than in the United States. European companies have sometimes drilled through shale to reach other rock formations, but they have rarely taken detailed measurements or collected samples of the shale layers.

So far, Poland's shale formations have attracted the most attention within the region. The nation depends heavily on coal, and what natural gas it does use comes almost exclusively from Russia. In the mid-2000s, the burgeoning US shale boom prompted Poland's government to offer shale exploration licences that went to local companies as well as major international energy firms, including the US companies ExxonMobil and Chevron, and the French firm Total. Poland's foreign minister, Radosław Sikorski, said in 2010 that Poland would become "a second Norway" — referring to Europe's second-largest natural-gas producer, after Russia.

The excitement was bolstered in 2011 by an assessment from Advanced Resources International (ARI), a consultancy in Washington DC that was commissioned by the US Department of Energy to study shale-gas resources worldwide. That study estimated the quantity of shale rock and other parameters such as the total organic content of the rock, which is the source of oil and gas. ARI also estimated parameters to represent the risk that some shale zones, or plays, might not prove promising or that only a portion of them might be amenable to drilling. Given these assumptions, ARI calculated that Poland's shale-gas plays hold about 5,295 billion cubic metres (bcm) of technically recoverable gas, the most shale gas of any nation in Europe. If all of that gas could be extracted, it

would be equivalent to 325 years of Poland's current gas consumption<sup>1</sup>.

While companies began drilling dozens of test wells in Poland, the Polish Geological Institute (PGI) in Warsaw made its own estimate in March 2012. Taking the considerable uncertainty over the data into account, the PGI calculated that Poland has 346–768 bcm of recoverable shale gas onshore — about one-tenth of ARI's figure<sup>2</sup>.

Then in July 2012, the US Geological Survey (USGS) released another study of Poland's shale-gas resources. The agency assumed that

# "IT IS TOO SOON TO DISMISS POLAND'S EXTENSIVE SHALE POTENTIAL."

individual wells would yield about half as much gas as the PGI assumed and that the area that is likely to contain recoverable gas is only about one-third of the size. So the USGS wound up with an estimate even smaller than the other two, with a mean result of just 38 bcm of recoverable gas, and a huge range of uncertainty, from 0 to 116 bcm. The mean was about one-tenth that of the PGI's estimate, and about one-hundredth of ARI's<sup>3</sup>.

"One report — huge potential. A year later — nothing," says PGI geologist Hubert Kiersnowski. "The scale of uncertainty is so big."

Meanwhile, results started coming in from test wells. Of the 72 wells drilled by the end of 2015, 25 were successfully fracked to release gas. However, these wells yielded only about one-third to one-tenth of the flow that would be required to turn a profit, says petroleum geologist Paweł Poprawa of AGH University of Science and Technology in Krakow, Poland, and formerly of the PGI. None of the wells has become a commercial producer.

At the peak of interest in early 2013, companies held shale-drilling licences covering about one-third of Poland. But throughout 2013 and 2014, the major international energy firms gave up their shale-exploration licences and left the country, often citing disappointing results. The last to leave was Texas-based ConocoPhillips in June 2015 — now Poland's shale drilling is almost at a standstill.

One major hurdle to development is that Poland's shale is expensive to drill because it is

buried around 3–5 kilometres down, compared with around 1–2 kilometres for most successful US plays. Some of Poland's shale also has a high clay content, which makes the rock harder to fracture. And exploratory holes into one of Poland's most promising shale formations — in the north, near the Baltic Sea — showed that it held a geological barrier that would limit how much gas could be tapped by individual wells, says Poprawa. The drilling results suggest that ARI "overestimated the acreage, the thickness, and the quality of the shale," he says.

The PGI says that its previous lower estimates are reinforced by its latest, as-yet-unpublished assessment, which draws on recent shale-drilling tests. PGI spokesperson Andrzej Rudnicki calls ARI's much higher estimates "enthusiastic, but geologically unrealistic".

"The results in Poland to date indeed have been disappointing," concedes geologist Scott Stevens of ARI. He says that the main reason for the unproductive wells was "extremely high" stresses in the rock, which makes fracking less effective. "There was no way that the exploration companies could know that in advance," he notes. Nonetheless, he argues, "It is too soon to dismiss Poland's extensive shale potential." Given the limited available data, he does not see a reason to revise ARI's estimate.

Even the PGI's lower estimates suggest that there is a still a substantial amount of gas trapped in Poland's shale. However, it is uncertain whether any of that gas will be profitable to extract. "I am still hopeful," Poprawa says. "But the initial hopes were not realistic."

## DASH FOR GAS

Although companies raced to grab concessions in Poland, activity in the United Kingdom has been subdued. In 2011, Cuadrilla Resources fracked the United Kingdom's first shale well near Blackpool in northern England, but this triggered two small earthquakes, which led the government to place a year-long moratorium on further fracking. After the moratorium lifted, companies slowly began vying to tap UK shale.

According to a 2013 assessment by ARI, UK shale holds 17,600 bcm of gas. Only 728 bcm of this is judged to be technically recoverable: if that could be profitably extracted, it would satisfy the United Kingdom's gas needs for about a decade<sup>4</sup>.

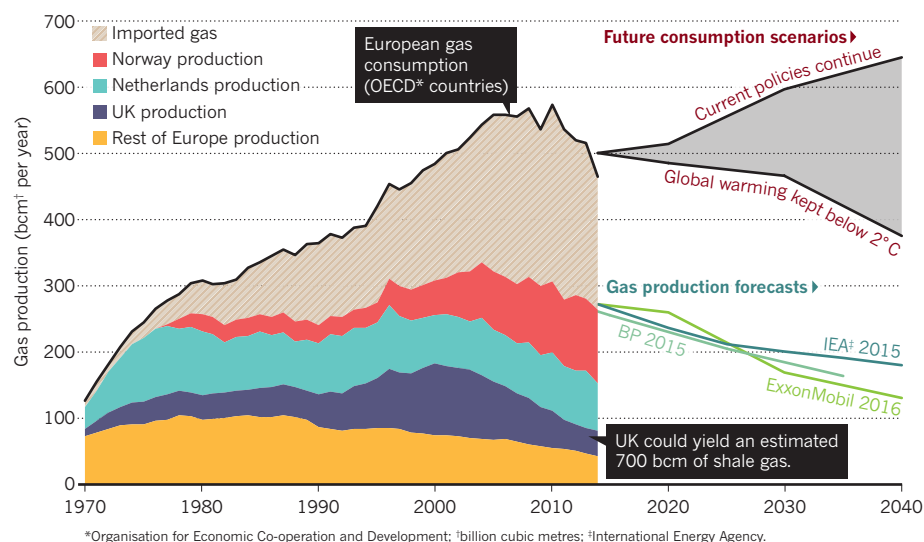
The British Geological Survey (BGS) has assessed the shale-gas resources in the United Kingdom's three major plays by constructing a 3D model of the subsurface using drilling records and seismic surveys, which has allowed it to roughly estimate the volume of shale rock. But geologist Ian Andrews of the BGS insists that this estimate is just a first pass based on the seismic information available, "which is sparse, and fairly poor".

By testing old rock cores stored by the government, the BGS was also able to measure some of the properties of UK shale, such as the total organic carbon (TOC) content. Successful shale plays in the United States typically



## LOOMING GAS CRUNCH?

Domestic production of natural gas has fallen in Europe (except Russia) and forecasts predict further declines. That could force nations to increase their imports, depending on consumption trends. Fracking could unlock additional supplies, but drilling results so far have not been promising.



have TOC values greater than 2%. Although TOC measurements for the United Kingdom are scant, the available data suggest that there are large volumes of rock above the 2% threshold. But data are lacking for other key parameters, such as the rock porosity, which adds greatly to the uncertainty of these projections.

The BGS estimated that the three shale plays it has assessed so far hold around 39,900 bcm of gas, with an uncertainty range of 24,700–68,400 bcm (refs 5,6). This is more than the ARI estimate, but that study only considered the most promising rock. The BGS did not attempt to estimate how much of that gas would be technically recoverable. “How much we can get out of the ground, I don’t think anybody knows yet, because the drilling hasn’t happened to test it,” says Andrews.

Although the BGS’s studies used US shale plays as analogues for crucial parameters, the two nations have different geological histories. The United States has large deposits of shale that are not too thick and have been folded little over time. The shale in the United Kingdom is more complicated, says petroleum geoscientist Andrew Aplin of the University of Durham, UK. “It’s been screwed around with more”, creating more folds and faults.

That greater complexity could pose challenges. One risk is that pumping fluid into rock can trigger earthquakes if the wells are near faults or large natural fractures. “It’s better to stay away from them, especially when they’re located near densely populated areas,” says natural-gas expert Rene Peters of the Netherlands Organisation for Applied Scientific Research (TNO) in the Hague. But there has been relatively little high-resolution seismic imaging in Europe, he says, so “not all these fractures are known”. Small faults can pose another challenge. If the fracking fluid leaks into a fault, the pressure on the rock is reduced

and the fracking is less effective. Given the geological hurdles and the United Kingdom’s dense population, it may prove difficult to find many promising, acceptable places to drill.

### FAR FROM PROFITABILITY

The United Kingdom’s appetite for gas is expected to grow sharply. In November, the government set out the goal of phasing out coal-fired power plants by 2025, unless they have carbon capture and storage systems. The government expects nuclear, wind and solar power to play a part in filling the void left by coal — but natural gas would be the linchpin because it produces less carbon dioxide and other pollution than does coal, and existing infrastructure can be used to produce electricity from gas. “We’ll only proceed if we’re confident that the shift to new gas can be achieved within these timescales,” UK energy secretary Amber Rudd said in a speech announcing the policy shift. “We currently import around half of our gas needs, but by 2030 that could be as high as 75%. That’s why we’re encouraging investment in our shale-gas exploration so we can add new sources of home-grown supply.”

Other European nations are also counting on natural gas to help them to cut their coal use and meet their commitments under the United Nations climate treaty signed in Paris in December. But shale gas may not provide the answer. At the June 2015 World Gas Conference in Paris, industry speakers were pessimistic that Europe would see a fracking boom like that in the United States. Philippe Charlez, manager of unconventional resources development at Total, said that given the current costs for shale wells, “we are very, very far in Europe from profitability”.

Many assessments in the past two years — including those by the International Energy Agency and oil giants BP and ExxonMobil

— agree that Europe is unlikely to produce much shale gas, and that conventional gas production will continue to decline<sup>7–9</sup>. And if gas imports cannot make up the difference, says Stern, “Europe is going to have even more difficulty reducing carbon emissions”.

The most recent signs are not good for shale across the continent. Besides retreating from Poland, major petroleum companies have pulled out of nascent shale drilling efforts in Romania, Lithuania and Denmark, usually citing disappointing yields. Various members of the European Union from Bulgaria to France have instituted moratoria or bans on fracking, as have Scotland, Wales and Northern Ireland, all citing environmental concerns.

England is home to some of the few remaining attempts to tap shale gas in Europe. A handful of companies have applied for permission to drill, which could finally reveal whether the United Kingdom’s shale deposits will be a jackpot or a dud. But environmentalists have put up a strong fight, and permissions have been slow to emerge.

Cuadrilla requested approval in January 2015 to drill beneath the undulating fields of Lancashire, but the county council rejected the request in June over concerns about traffic, noise and the visual impact of drilling. That decision and the broader difficulties that confront fracking in Europe leave the future of natural gas there in limbo. To figure out whether any play has potential, companies must drill as many as 50 to 100 wells. But the public opposition and the poor drilling results so far mean that companies are not eager to sink that kind of effort into fracking in Europe right now, says Stern. “I can’t see any country, including the UK, where that will happen anytime soon.” ■

**Mason Inman** is a reporter in Oakland, California. Travel for this article was supported by the European Geosciences Union’s Science Journalism Fellowship.

1. US Energy Information Administration. *World Shale Gas Resources* (EIA, 2011); available at <http://go.nature.com/yq9cbo>
2. Polish Geological Institute. *Assessment of Shale Gas and Shale Oil Resources of the Lower Paleozoic Baltic-Podlasie-Lublin Basin in Poland — First Report* (PGI, 2012); available at <http://go.nature.com/buxvqp>
3. US Geological Survey. *Potential for Technically Recoverable Unconventional Gas and Oil Resources in the Polish-Ukrainian Foredeep, Poland, 2012* (USGS, 2012).
4. US Energy Information Administration. *Technically Recoverable Shale Oil and Shale Gas Resources* (EIA, 2013); available at <http://go.nature.com/lkxmxe>
5. British Geological Survey. *The Carboniferous Bowland Shale Gas Study: Geology and Resource Estimation* (BGI, 2013); available at <http://go.nature.com/fdavyi>
6. British Geological Survey. *The Carboniferous Shales of the Midland Valley of Scotland: Geology and Resource Estimation* (BGI, 2014); available at: <http://go.nature.com/dbdytt>
7. International Energy Agency. *World Energy Outlook 2015* (IEA, 2015).
8. *BP Energy Outlook 2035* (BP, 2015); available at <http://go.nature.com/s1zomr>
9. *The Outlook for Energy: A View to 2040* (ExxonMobil, 2016); available at <http://go.nature.com/fkpiif>

SOURCES: BP/EIA/IEA/EXXONMOBIL

*Virus-sized particles that fluoresce in every colour could revolutionize applications from television displays to cancer treatment.*

# THE NANOSCALE RAINBOW



BY XIAOZHI LIM

**A**t Biopolis, a sprawling research complex in Singapore, Chi Ching Goh leans over an anaesthetized mouse lying on the table in front of her, and carefully injects it with a bright yellow solution. She then gently positions the mouse's ear underneath a microscope, and flips a switch to bathe the ear in ultraviolet light. Seen through the microscope's eyepiece, the illumination makes the blood underneath the skin glow green, tracing the delicate vessels that carry the solution through the creature's body.

Ultimately, Goh, a PhD candidate at the National University of Singapore, hopes that the method will help her to find blood vessels that are leaking owing to inflammation, perhaps helping to detect malaria or predict strokes. Crucial to her technique are the virus-sized particles that give the solution its colour. Just a few tens of nanometres across, they are among

a growing array of 'nanolights' that researchers are tailoring to specific types of fluorescence: the ability to absorb light at one wavelength and re-emit it at another.

Many naturally occurring compounds can do this, from jellyfish proteins to some rare-earth compounds. But nanolights tend to be much more stable, versatile and easier to prepare — which makes them attractive for users in both industry and academia.

The best-established examples are quantum dots: tiny flecks of semiconductor that are prized for their beautiful, crisp colours. Now, however, other types of nanolight are on the rise. Some have a rare ability to absorb lots of low-energy photons and combine the energy into a handful of high-energy photons — a trick that opens up opportunities such as producing multiple colours at once. Others are made from

polymers or small organic molecules. These are less toxic than quantum dots and often outshine them — much to the amazement of chemists, who are used to carbon-based compounds simply degrading in the presence of ultraviolet light.

"I was kind of surprised to find that we can make organic particles much brighter than inorganic particles," says Bin Liu, a chemical engineer at the National University of Singapore and the designer of the fluorescent nanoparticles that Goh is using.

Nanolights have already begun to find application in areas ranging from flat-screen displays to biochemical tests. And researchers are working towards even more ambitious uses in fields such as solar energy, DNA mapping, motion sensing and even surgery. "The research is certainly fast-paced," says Daniel Chiu, who studies fluorescent nanoparticles

**Solutions of quantum dots made from cadmium selenide respond to ultraviolet light by emitting visible light in specific wavelengths.**

at the University of Washington in Seattle.

It is also increasingly wide ranging, adds Paul Alivisatos, a chemist at the University of California, Berkeley, and a co-founder of the first quantum-dot technology companies. “It’s so much fun now.”

## SIZE MATTERS

The nanolight era began with the discovery of quantum dots in 1981. Russian physicists were growing tiny crystals of the semiconductor cuprous chloride in silicate glass and observed that the colour of the glass depended on the size of the particles<sup>1</sup>. The crystals were so small that quantum effects were kicking in and they were behaving somewhat like atoms: they could absorb or emit light only as specific colours, with the exact frequencies depending on the size or shape of the particles (see ‘Bridge the gap’).

The quantum dots were bright and beautiful, says Yin Thai Chan, who studies them at the National University of Singapore, but “there were no obvious applications”. By the early 2000s, however, the pure colours had begun to attract television manufacturers, as well as biomedical researchers, who saw their potential for labelling specific proteins and DNA segments.

“Everything is good about quantum dots,” says Liu — except for one thing: their toxicity. The best-performing dots contain cadmium, which can poison cells. This limits their usefulness in biology and in applications such as household electronics, because some countries do not allow use of the element in such devices. To some extent, this problem can be overcome by replacing cadmium with zinc or indium, which are considerably less toxic, or by wrapping cadmium-based quantum dots in polymers that are biocompatible. But the toxicity is still a drawback for researchers who are pursuing ambitious applications such as fluorescence-guided surgery, in which nanoparticles are injected into a tumour, for instance, to make it glow and help surgeons to remove all traces of it.

## GOING ORGANIC

Partly in response to this challenge, researchers have begun to develop nanoparticles from materials that fluoresce naturally. Because the light-emitting properties of these nanolights come from their composition rather than their size or shape, they are easier to make with specific colours. “Practically, this is useful because of the difficulties to synthesize everything in the same size,” says Chiu.

It also frees up nanolight researchers to explore alternative materials, such as semiconducting polymers. Studied for their potential in electronics since the 1950s, these polymers consist of simple compounds linked into a long chain in which electrons are free to move, but only at certain energies

determined by the chain’s composition.

Light is emitted when electrons are kicked up to higher energy levels by some outside source, such as ultraviolet light, then fall back down to lower levels. The polymers can also be decorated with side groups to give them specific properties — for example, targeting them to cancer cells, or helping them to dissolve in water. And when chains are aggregated into polymer nanoparticles, or ‘P-dots’, they can be as much as 30 times brighter than a quantum dot of comparable size<sup>2</sup>.

Semiconducting polymers do tend to be less stable than the inorganic semiconductors used in quantum dots. But because they are based on carbon, and contain no metals, they are much more likely to be biocompatible. P-dots have been used to stain and image cells, and also as sensors to detect oxygen, enzymes or metal ions such as copper.

In 2013, for example, Chiu and his collabora-

tors reported that a P-dot bound to a terbium ion can detect biomolecules produced by bacterial spores<sup>3</sup>. Under an ultraviolet lamp, the P-dots glow dark blue and the terbium ions emit a faint neon green colour. But when passing biomolecules attach themselves to terbium, the ions’ light strengthens to a bright green. The P-dots’ light remains unchanged, so it serves as an internal standard.

Unfortunately, P-dots also have a fundamental problem: the polymer molecules are packed together so closely that they can be affected by ‘quenching’ — a phenomenon in which most of the energy coming from the original light source is quickly dissipated and fails to trigger fluorescence.

Quenching has a huge impact on efficiency, says Yang-Hsiang Chan, a chemist at National Sun Yat-Sen University in Kaohsiung, Taiwan. One way to tackle it is to add bulky groups onto the polymer backbone to prevent the polymers from getting too close to each other. But this can be self-defeating: the resulting nanoparticles tend to be too fat to get into cells, say, or too dim to be useful. “It is very hard to get the right balance,” says Chan, who is working to solve the problem by designing new polymers.

## TOGETHER WE SHINE

A more fundamental solution was pioneered in 2001, when Ben Zhong Tang at the Hong Kong University of Science and Technology in Clear Water Bay found that a class of small organic molecules would fluoresce only when they aggregate together<sup>4</sup>. These molecules are shaped like propellers or pinwheels, and they fluoresce when packed because they can no longer move and waste their energy. Instead, they release their energy as light — a phenomenon Tang has named aggregation-induced emission (AIE).

He called the molecules AIE-gens.

Over the next few years, Tang and his students changed the side groups and introduced elements such as nitrogen or oxygen, and AIE-gens can now glow in the entire spectrum of colours from ultraviolet to near-infrared. “My students quickly made a lot,” says Tang. “We can change the colour at will.”

In 2011, Tang met Liu through a collaboration at the Institute of Materials Research and Engineering in Singapore, part of the government-backed Agency for Science, Technology and Research (A\*STAR). At that time, AIE-gens were performing well, except that they could not dissolve in water, which made them difficult to use in biological applications. Liu was an expert in making things water-soluble, so Tang gave her some of his best AIE-gens to work with.

Liu solved the problem by experimenting with polymers that are oil-loving on one end and water-loving on the other. The AIE-gens

# “WE CAN CHANGE THE COLOUR AT WILL.”

crowd within the polymer’s oil-loving ends, and its water-loving ends point outwards to form a protective shell, resulting in a water-soluble capsule with a dense core full of AIE-gens. Liu designed a protective shell for the resulting nanoparticles, called AIE-dots, such that it could be decorated with various chemical groups that are tailored to specific applications. The shell can easily accommodate a wide variety of AIE-gens, says Liu, “so that we can screen a lot of molecules very quickly to find out which one is the best.”

AIE-dots have been used to stain various tissues, from blood vessels to cancer cells to intracellular organelles such as mitochondria. Last year, Liu, Tang and their colleagues reported an AIE-dot that could be useful in a type of light-activated treatment known as photodynamic therapy<sup>5</sup>. It carries two molecules on its surface: one to get the dot into a cancer cell, and another to make it stick to the mitochondria. Once excited by an external light source, the AIE-dot produces red light that generates oxygen radicals near the mitochondria and kills the cancer cells.

The best AIE-dots can be 40 times brighter than quantum dots<sup>6</sup>. “With AIE, high density in constrained space produces high brightness,” says Guangxue Feng, a research assistant in Liu’s lab. That is particularly useful for applications such as visualization of tissues or long-term tracking of cancer cells, which halve the number of nanoparticles per cell every time they divide.

But the brightness comes at a cost: AIE-dots produce a much broader, more-muted spectrum than the pure, brilliant colours of quantum dots. But that hasn’t kept Liu from starting LuminiCell, a spin-off company in Singapore that produces AIE-dots in three colours and three sizes for research such as Goh’s at A\*STAR.



Tang is also trying to start a company; both he and Liu are now hoping to gain approval from the US Food and Drug Administration to test AIE-dots for human use in applications such as fluorescence-guided surgery.

## INTO THE INFRARED

Another thing that limits the biological use of nanolights is that most of them absorb ultraviolet or visible light, which can penetrate only a few millimetres into tissue. Longer-wavelength near-infrared radiation can penetrate up to three centimetres — a much better depth for uses such as releasing drugs. But infrared light does not have enough energy to break the bonds that hold drugs on the nanoparticle, so many researchers are turning to a process called upconversion. This involves making material that can absorb multiple low-energy infrared photons, accumulate the energy and then re-emit it as higher-energy ultraviolet or visible photons.

The group of heavy-metal elements known as lanthanides are particularly good at this trick. In 2011, Xiaogang Liu at the National University of Singapore reported that his laboratory had created a particularly versatile type of nanoparticle<sup>7</sup> with a Russian doll-like structure. It consists of a series of concentric shells that each contains a different combination of lanthanides. The energy from infrared light is absorbed by the core, then migrates outwards layer by layer, snowballing from lanthanide to lanthanide before finally emerging as high-energy light near the surface.

The 15 lanthanides can be combined in numerous different ways to produce nanoparticles that emit in all colours, sometimes even several at once. In one demonstration, a student in Liu's lab shone an infrared laser through a series of beakers containing clear solutions of the nanoparticles: glowing lines of purple and green light appeared in the beakers where the infrared beam passed through.

Liu thinks that these upconversion nanoparticles have tremendous potential in photovoltaics, where they could help to capture near-infrared light, which makes up almost half of the Sun's radiation. This is a long way from being practical, however: the brightest available nanoparticles convert just 10% of the light they absorb. Liu's group is working to build a library of these nanoparticles — no small task considering the number of lanthanides — to systematically study their properties and work on making them brighter.

Last December, Marta Cerruti, a biomaterials scientist at McGill University in Montreal, Canada, reported a proof-of-concept system in which a lanthanide-containing nanoparticle is coated with a gel that contains a 'drug' — for testing purposes, a compact, stable protein<sup>8</sup>. After absorbing near-infrared light, the nanoparticle emits infrared, visible and ultraviolet light simultaneously. The infrared emission allows the researchers to track the nanoparticle's location, and the ultraviolet light cleaves the protein's bond to the gel and releases it — or at least, it has in the laboratory. Cerruti's group

is now planning tests in animals.

At the end of the day, quantum dots are still the nanolights to beat. "They are the de facto standard," says Chan. "A lot of the fundamental phenomena concerning light emission are established in quantum dots and it shapes the way others explain what they see."

Quantum dots are also still a research frontier. For example, they are getting a boost from relatively new semiconducting materials such as the perovskites. Unlike conventional semiconductors, which have a fixed ratio of elements, perovskites can have variable ratios, so researchers can tailor the dots' emission by varying their composition as well as their size. "They have two degrees of freedom for tunability," says Edward Sargent, a materials engineer at the University of Toronto, Canada.

Last year, Sargent reported a hybrid material in which quantum dots are held within a perovskite<sup>9</sup>, yielding the kind of high brightness and good electron mobility that manufacturers might like for use in flat-screen displays.

Other researchers are hoping to combine the best properties of each component by pursuing hybrid nanolights. Bin Liu, for example, is trying to blend AIE-dots with quantum dots to produce narrow emissions. And semiconducting polymers paired with AIE-dots can produce much brighter particles than either alone<sup>10</sup>.

Another grand challenge for nanolights is to create versions that emit infrared wavelengths efficiently. That would open up applications in motion sensing, from tiny detectors that tell the screen to turn off when a mobile phone is lifted to the ear to sophisticated devices for self-driving cars and home monitoring for elderly people. "There's so much more we could do," says Sargent. ■

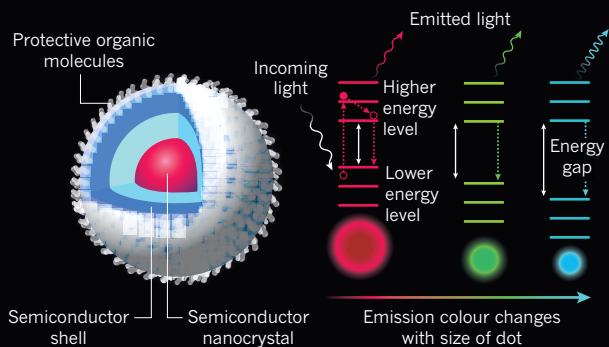
**XiaoZhi Lim** is a freelance writer in Singapore.

## BRIDGE THE GAP

Some virus-sized particles can be tailored to absorb light and fluoresce at specific frequencies. Light kicks electrons up to higher energy levels, and the glow is emitted when the electrons fall back past a wide energy gap.

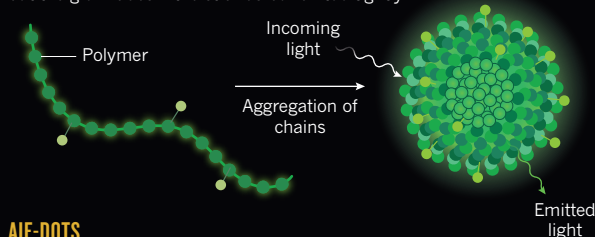
### QUANTUM DOTS

These particles contain layers of crystalline semiconductor material. The colour they emit depends on the particle's size.



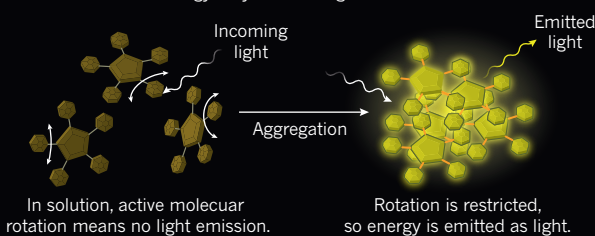
### P-DOTS

These particles are made of semiconducting polymers that emit light when they aggregate. They can be brighter than a quantum dot of the same size, but the glow fades if the strands bunch too tightly.



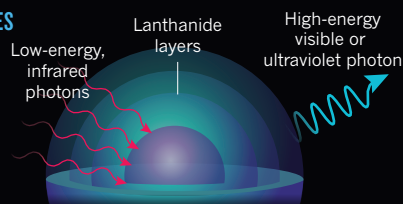
### AIE-DOTS

These molecules will not fluoresce unless they are tightly packed, which restricts the energy they lose through internal motion.



### UPCONVERSION PARTICLES

Particles made from layers of heavy-metal lanthanide elements can accumulate energy from infrared light and re-emit it as visible or ultraviolet light.



- Ekimov, A. I. & Onushchenko, A. A. *JETP Lett.* **34**, 345–349 (1981).
- Wu, C. *et al. J. Am. Chem. Soc.* **132**, 15410–15417 (2010).
- Li, Q. *et al. Anal. Chem.* **85**, 9087–9091 (2013).
- Luo, J. *et al. Chem. Commun.* **2001**, 1740–1741 (2001).
- Feng, G., Qin, W., Hu, Q., Tang, B. Z. & Liu, B. *Adv. Healthcare Mater.* **4**, 2667–2676 (2015).
- Li, K. *et al. Sci. Rep.* **3**, 1150 (2013).
- Wang, F. *et al. Nature Mater.* **10**, 968–973 (2011).
- Jalani, G. *et al. J. Am. Chem. Soc.* **138**, 1078–1083 (2016).
- Ning, Z. *et al. Nature* **523**, 324–328 (2015).
- Ding, D. *et al. Adv. Healthcare Mater.* **2**, 500–507 (2013).



# COMMENT

**HEALTH** The rich legacy of tracing thousands of lives over many decades **p.32**



**ECOLOGY** All of biology distilled into six, predictive rules **p.34**

**EQUALITY** US National Institutes of Health to crack down on sexual harassment **p.35**

**LAW** Australians reject change to primate regulations **p.35**

ARKADIUSZ PODNIESIŃSKI



An abandoned supermarket in the Fukushima prefecture in July 2015.

## Five years on from Fukushima

To build sustainability and trust, energy and environment research in Japan must become more interdisciplinary and global, say **Masahiro Sugiyama** and colleagues.

**N**ext week will mark five years since 11 March 2011, the day of the devastating Tohoku earthquake and tsunami, and the accident that followed at the Fukushima Daiichi nuclear-power plant of the Tokyo Electric Power Company. The quake and tsunamis killed nearly 16,000 people and injured more than 6,000; 2,600 are still missing.

Fukushima prefecture, the location of the crippled nuclear plant, was hit particularly

hard<sup>1</sup>. The Japanese government provisionally rated the severity of the accident on par with the 1986 Chernobyl disaster — a seven on the seven-point International Nuclear and Radiological Event Scale<sup>2</sup>. Around 110,000 people had to evacuate because of dispersed radionuclides. Despite the large-scale decontamination efforts, about 70,000 former residents are yet to return.

Shocked by the fallout, Japan changed its energy policy. The year before the disaster,

the 2010 Basic Energy Plan had called for 53% of electricity generation to come from nuclear power by 2030, implying significant new construction. Since the accident, Japan's energy policy has featured an expanded role for renewables and market liberalization — transitioning from a regional-monopoly model to one that is open to competition. (Some policy changes were made after the change in government in December 2012.)

In July 2012, reflecting the public's ►

► desire for a transition towards renewable energy, Japan introduced a feed-in tariff, which guarantees renewable-energy generators a high price for the electricity that they feed into the grid; it was particularly generous for solar photovoltaics. Installed solar capacity more than quadrupled in the first three years. A whopping 80-gigawatt capacity — 90% greater than Japan's nuclear-power capacity — has been approved. Next month, the retail electricity market will be fully liberalized, attracting a huge number of new entrants to the previously closed market.

The government also strengthened nuclear security substantially. A newly created, independent Nuclear Regulation Authority instituted new safety measures in 2013. All of Japan's nuclear-power plants halted operation for inspection after the accident, and the share of nuclear in power generation dropped from about 30% in 2010 to zero in 2014. Only 4 of the 44 reactors have been restarted so far.

The public debate on nuclear rumbles on. Still an important resource for this resource-poor nation, it is expected to provide 20–22% of electricity by 2030, under a new long-term energy strategy created in conjunction with Japan's pledge to the United Nations to cut greenhouse-gas emissions.

## WEAK LINKS

The journey since 2011 has been difficult, with policy controversies on every front. Many of the fraught decisions — on evacuation, clean-up, energy transition and disaster preparedness — were at the science–policy interface<sup>3</sup>. And scientists, especially those involved in giving policy advice, lost credibility and the trust of the public<sup>4</sup>.

Several initiatives have been launched to rebuild these crucial bridges. One such is an effort by the Japan Science and Technology Agency (JST) on research into scientific advice, and a 'deliberative polling' exercise in August 2012 involving the public that was used to inform the energy policy of the previous government. These efforts are yielding genuine progress, but slowly.

We strongly believe that the events and aftermath of 11 March highlighted a fundamental problem with research in Japan: weak connections between disciplines and between Japan's scholars and those working in other countries. In a nation that performs world-class research in conventional disciplines<sup>5</sup>, interdisciplinary scholarship lags, and Japanese researchers are keenly aware of this. Moreover, the nation's breadth of disciplinary coverage is narrower<sup>6</sup> and the rate of international collaboration is lower than in comparable nations (see 'Build bridges').

This has a particularly important implication for energy and environment research, which require the integration of diverse knowledge<sup>6–7</sup> that can come from anywhere in the world. During the Fukushima crisis,

researchers who were not used to collaborating with other disciplines (or other nations) struggled to do so<sup>8</sup>.

## TWO CASE STUDIES

Two examples illustrate the problem. The first concerns assessing the safety of nuclear-power plants. Probabilistic risk assessment (PRA) is a standard tool to quantitatively evaluate the likelihood of severe accidents and their impacts, using analysis methods such as fault trees and Monte Carlo simulation. Before the earthquake, nuclear experts conducting PRA research in Japan focused on internal events at nuclear plants (mechanical component failures and human errors), dealing mostly with engineering knowledge.

What the disaster vividly demonstrated is that nuclear plants are susceptible to external events and that accident impacts are not contained — they may include the release of radionuclides, with dire environmental effects. The PRA has therefore been extended beyond nuclear engineering to cover disciplines ranging from seismology and geology to atmospheric science and ecological modelling.

Before 2011, such interdisciplinary PRA research in Japan was limited compared with other developed economies that have significant nuclear presence such as in the United States, the United Kingdom and France (see

go.nature.com/sev71o; in Japanese). This was partly because the country did not require PRA for regulatory purposes. Under the new 2013 regulations, Japan mandated the use of PRA for nuclear plants and is now trying to catch up in this area.

The second example concerns innovation in renewable energy. Although many citizens would prefer the nation's energy portfolio to have a larger share of renewables, these are more costly in Japan than elsewhere, even for technologies such as solar photovoltaics, in which Japan was a pioneer.

Ideally, Japan should explore how each policy alternative might affect future cost trends for solar. Combining energy-systems analysis and policy analysis with technology forecasting methods (based on expert elicitation, bottom-up engineering analysis and learning curves that describe empirical relationships between cumulative production and cost reduction) would yield crucial insights. Such interdisciplinary studies in Japan are hard to come by. Because of the low level of global networking in these areas, international experiences are not widely appreciated in Japan either.

This has already affected the solar market in Japan. Under the feed-in tariff, developers rushed to install expensive solar devices, costing consumers trillions of yen that could have been saved by gradual installations made in tandem with cost reductions. In Germany, by contrast, there was a clear incentive for solar developers to reduce cost under its fine-tuned tariff scheme with frequent price adjustments.

Our critics will say that these issues are political, not academic. We feel that this attitude is the source of the problem. Engaged scholarship is a prerequisite for informed policymaking. Scientists and social scientists must do their part.

## TWO FIXES

Two big changes would go a long way to improving interdisciplinary research in energy and the environment in Japan. Going global is the key, and will pay dividends: Japan would leverage international expertise, and the rest of the world would learn from Japan's experiences.

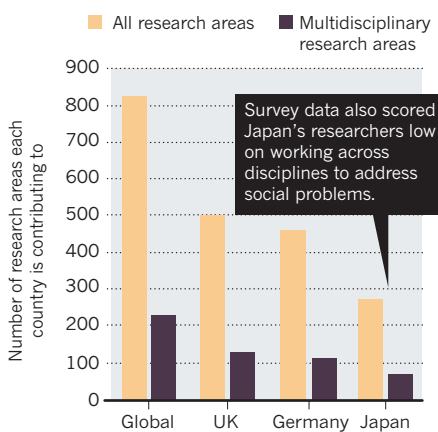
**Globalize the review process.** Because of the small number of researchers engaged in interdisciplinary research, the pool of reviewers for academic journals and funding proposals is limited. In policy-relevant interdisciplinary research, particularly in energy and environment, publishers and granting programmes, such as the government-backed KAKENHI (Grants-in-Aid for Scientific Research), should make parts or all of their review processes international. The connections made could also boost international collaboration.

For strategic research in energy and the environment, funding agencies should

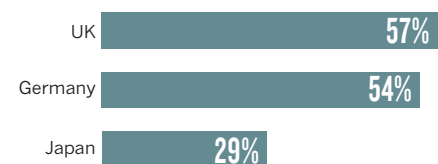
## BUILD BRIDGES

The breadth of disciplinary coverage is narrower in Japan than in nations that have similar research systems, and international collaboration is lower.

### 1 Contribution by country to research areas publishing the most highly cited papers (top 1% by citation, 2007–12)



### 2 Share of papers with international co-authors (2011–13)



SOURCES: REF. 5 (1); A. SAKA & M. IGAMI BENCHMARKING SCIENTIFIC RESEARCH 2015 (NISTEP, 2015) (2); 2014 NISTEP EXPERT SURVEY ON JAPANESE S&T AND INNOVATION SYSTEM (CALLOUT)





Employees, and thousands of paper cranes, inside the Fukushima plant three years after the accident.

require scientists to publish part of their results in international journals even for policy-oriented research, whose target readership obviously prefers Japanese to English. Many papers, although tailored to the policy context of Japan, would appeal to global experts because energy and environmental issues are global. Large-scale programmes of the science-and-technology ministry, and strategic research funds of the environment ministry, should take the lead. Because policymakers need deliverables to be communicated in Japanese as well, this will increase the burden on researchers, which should be reflected in their funds.

In April, Japan will start a new round of its Science and Technology Basic Plan, a cabinet-level, five-year policy on research and innovation. It is good to see that building an international researcher network is one of the key agenda items. Japan must make that vision into a reality.

**Globalize research.** Strategic, policy-oriented research programmes in Japan should be designed so that they can benefit from international experience and domestic experience can be shared globally. For example, the Collaborative Laboratories for Advanced Decommissioning Science (CLADS), established as a research base for the decommissioning of the Fukushima plant, should be more outward-facing.

Decommissioning involves many disciplines, including nuclear engineering, meteorology and oceanic-risk assessments, ecology and remediation. By soliciting

international research proposals, CLADS should involve more researchers from elsewhere in Asia, where many countries have nuclear ambitions, including China, South Korea, India and many southeast Asian countries. Working with overseas scientists, CLADS should publish some outcomes in English.

Another opportunity is Future Earth, a ten-year global sustainability research initiative that puts interdisciplinarity at the forefront

**“Going global is the key, and will pay dividends.”**

alongside stakeholder engagement<sup>9</sup>. For its contribution, Japan should elevate energy research to a key component. Japan has several advanced energy technologies, but to move them into the market at scale requires outside input, particularly when it comes to innovation policy.

As Asia becomes the centre of the global energy economy, the time is ripe for Japan, as part of the Future Earth platform, to embark on a truly interdisciplinary and international project, and colleagues from neighbouring nations should do the same. Such initiative should receive rigorous academic oversight from an international advisory body.

#### BETTER TOGETHER

This year is also the 30th anniversary of the Chernobyl nuclear accident. In Europe, and Germany in particular, that disaster spawned fresh thinking on many fronts. The German book *Risk Society* by Ulrich Beck, published in 1986 soon after the accident, explored how

risks from technology and industrialization shape modern society.

The disaster catalysed a transition away from nuclear to renewables, which is now gathering renewed momentum, backed up by interdisciplinary studies on energy transformation. As in Germany in the late 1980s, Japan has seen many fresh attempts to carve out new directions for research, but so far such efforts have been fragmented and scattered, many along disciplinary lines.

Five years on from March 2011, problems abound. Fukushima and the Tohoku areas are yet to recover, and the transition towards renewables has been rocky. Most, if not all, of these issues are fundamentally political and socio-economic<sup>3</sup>. But scientists, social scientists and their funders must engage. Without better connections across disciplines and nations, the science-policy interface cannot improve. The people of Japan deserve better. ■

**Masahiro Sugiyama** is assistant professor at the Policy Alternatives Research Institute, University of Tokyo, Japan. **Ichiro Sakata** is director of the Policy Alternatives Research Institute and professor in the Graduate School of Engineering, University of Tokyo, Japan. **Hideaki Shiroyama** is dean of the Graduate School of Public Policy and professor in the Graduate Schools for Law and Politics, University of Tokyo, Japan. **Hisashi Yoshikawa** is a project professor at the Graduate School of Public Policy and Policy Alternatives Research Institute, University of Tokyo, Japan. **Taketoshi Taniguchi** is professor in the Policy Alternatives Research Institute, University of Tokyo, Japan. e-mail: masahiro\_sugiyama@alum.mit.edu

1. Ofuji, K. *WIREs Energy Environ.* **4**, 307–315 (2015).
2. International Atomic Energy Agency Re-evaluation of INES rating: Effect to the Nuclear Facilities from the Earthquake on East Area of Japan (IAEA, 2011); available at [go.nature.com/jqtejr](http://go.nature.com/jqtejr)
3. Fujigaki, Y. (ed.) *Lessons from Fukushima: Japanese Case Studies on Science, Technology, and Society* (Springer, 2015).
4. Arimoto, T., & Sato, Y. *Science* **337**, 1176–1177 (2012).
5. Igami, M. & Saka, A. *Scientometrics* **106**, 383–403 (2016).
6. Yoshikawa, H. *Trends Sci.* **7**, 25–34 (2002); available at <http://doi.org/bcct4> (in Japanese).
7. Kajikawa, Y. *Sustain. Sci.* **3**, 215–239 (2008).
8. Shiroyama, H. (ed.) *The Fukushima Daiichi Nuclear Power Plant Accident and Complex Risk Governance* (Toyo Keizai, 2015).
9. Yasunari, T., Niles, D., Taniguchi, M., & Chen, D. *Curr. Opin. Environ. Sustain.* **5**, 288–292 (2013).

#### CORRECTION

The Comment ‘Current climate models are grossly misleading’ (*Nature* **530**, 407–409; 2016) gave the wrong citation for the review about IAMs use in climate economics. The correct reference is J. D. Farmer *et al.* *Environ. Res. Econ.* **62**, 329–357 (2015).



Baby boom: new arrivals at a London hospital in 1945.

## PUBLIC HEALTH

# Tracing the social roots of health

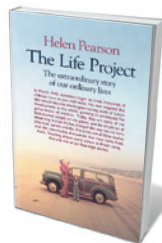
**Andrew Steptoe** applauds a cogent exploration of Britain's groundbreaking longitudinal birth-cohort studies.

**H**elen Pearson's fascinating *The Life Project* examines the history and legacy of the British birth-cohort studies, the first of which launched 70 years ago. These longitudinal investigations follow large samples of people 'recruited' as infants, periodically gathering data relevant to social and psychological development, schooling, employment and mental and physical health.

These studies — of, collectively, some 70,000 people — have played a crucial part in identifying how socio-economic circumstances drive inequalities in health and development. They have informed health, education and social policy, and provided a template for the 'life course' approach to health and development, studying how early experiences shape later outcomes. They are the envy of the world, and have contributed to topics as diverse as the perinatal determinants

of adult health, the establishment of free nursery places for British 3- to 4-year-olds and the drive to promote adult literacy and numeracy.

Pearson, *Nature's* chief features editor, reveals that the studies emerged haphazardly. The first — the National Study of Health and Development (NSHD), following 5,362 people born in England, Scotland or Wales in 1946 — was stimulated by 1930s concerns about the falling birth rate, but began, ironically, at the start of the post-Second World War baby boom. The National Child



**The Life Project: The Extraordinary Story of Our Ordinary Lives**  
HELEN PEARSON  
Allen Lane: 2016.

Development Study (NCDS), following 17,000 people born in 1958, aimed to address perinatal mortality, and collected data on deaths as well as live births. Perinatal health was also a major issue for the British Cohort Study (BCS70), following 17,000 born in 1970. The Millennium Cohort Study (MCS) follows 19,000 people born in 2000–02. Pearson also covers the local Avon Longitudinal Study of Parents and Children (ALSPAC) in southwest England — which focuses on 14,000 women recruited while pregnant in 1991–92, and their children. In addition, she looks at the Life Study, which was planned to begin in 2012 but never got off the ground.

Along with providing key information about maternal and neonatal health, the studies evolved wider remit. James Douglas, first director of the 1946 study, set the standard by collecting data on social and economic conditions, father's occupation, diet, health, temperament and behaviour, as well as birth weight and infant health and survival. Each cohort developed its own identity, a product of the principal investigators' interests and the era in which it was launched. As the NSHD's cohort turns 70 this year, the study is embracing detailed measures of brain function, cardiovascular activity and physical capability. ALSPAC, begun by epidemiologist Jean Golding in the era of molecular medicine, included collection of DNA and other biological samples from the beginning. BCS70, by contrast, is collecting biomarkers for the first time this year, when its participants are 46.

The NCDS moved towards social issues such as education when formidable educational psychologist Mia Kellmer Pringle took the lead. Its findings on the enduring impact of social and economic adversity on child development have fed into national debates about social mobility and cycles of deprivation. The MCS, initially led by demographer Heather Joshi, is now being used to investigate emerging twenty-first-century issues, such as the growth of childhood obesity, parental involvement in learning and the impact of birth season on educational attainment.

Some of the studies have had a bumpy ride, as Pearson relates. The 1946 cohort is overseen by the UK Medical Research Council (MRC), and this ensured regular assessment throughout childhood. But contact was not maintained with the 1958 sample, so participants had to be traced from scratch for data sweeps at ages 7 and 11. The BCS70 launched with funding from 23 organizations, then was virtually appropriated by the ebullient but eccentric paediatrician Neville Butler. He was able to continue the study by cobbling together funding from diverse sources such as wealthy aristocrats, philanthropists, actors and celebrities. He designed many of the data sweeps himself — and published results sporadically. The study almost fell off the map until it was rescued by the Social Statistics



Research Unit at London's City University in 1996. Perhaps even more extraordinary was the MCS, a mainly political initiative by the government of then-prime minister Tony Blair, devised to celebrate the millennium. Funding was not agreed until early in 2000, so planning was rushed and the design far from ideal: when the participants were recruited, the babies were already nine months old.

Pearson concentrates on social history rather than research findings, although she does highlight impacts on social and educational policy and medical practice. This is wise, because the science is difficult to synthesize: thousands of papers have been written on the studies' biomedical, social, economic and educational ramifications. Pearson inserts stories of participants such as Steve Christmas of the 1958 cohort, who overcame the disadvantages of a limited education with hard work and determination. These provide vivid accounts of growing up in different decades, and the roles of educational opportunity and psychological outlook in shaping lives.

As Pearson explains, the studies have had to fight for funding, owing in part to a tension between the need to maintain a sequence of measurements to understand how people change as they age, and the need to generate new hypotheses and collect new data in every funding period. At one time or another, each investigation was at risk of disbanding.

Which brings us to the melancholy tale of the Life Study. Planned by paediatric epidemiologist Carol Dezateaux to begin in 2012 with

**"Findings on the enduring impact of social and economic adversity have fed into national debates."**

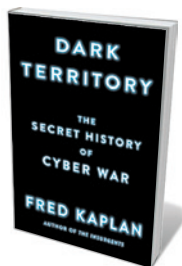
80,000 participants, the project aimed to take detailed measurements including data on air pollution, mobile-phone use and water quality; samples of blood, urine, saliva and faeces; and

recordings of behaviour. It was immensely complex, balancing the demands of medical and social scientists, and costs spiralled. Recruitment finally began early in 2015, but uptake was low, and the UK Research Councils withdrew funding within months.

It seems likely that there will be no new national birth cohorts any time soon. Yet it is encouraging that major funders such as the MRC and the Economic and Social Research Council have recognized the value of such studies. They are an important national research resource, and *The Life Project* does a great service in bringing them and the people at their heart to life for a general readership. ■

**Andrew Steptoe** is director of the Institute of Epidemiology and Health Care and British Heart Foundation professor of psychology at University College London. e-mail: a.steptoe@ucl.ac.uk

## Books in brief



### Dark Territory: The Secret History of Cyber War

Fred Kaplan SIMON AND SCHUSTER (2016)

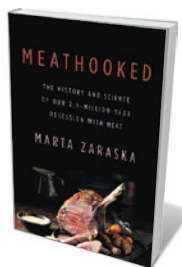
Beyond the bombs and drones of on-the-ground warfare, a shadow conflict is playing out in a virtual theatre of war. Covert cyberattacks — the hacking of digital systems that control everything from dams to centrifuges in uranium-enrichment facilities — are a new norm. Pulitzer-prizewinning journalist Fred Kaplan's taut, urgent history traces the dual trajectory of digital surveillance and intervention, and high-level US policy from the 1980s on. In 2014 alone, he reveals, almost 80,000 US security breaches occurred — an artefact of the very network connectivity that enriches the country's economy.



### Being a Beast

Charles Foster PROFILE (2016)

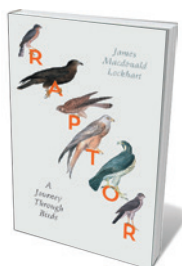
Humanity's obsession with animals spans shamanic 'possession' by wolves and a hefty tranche of children's literature. Charles Foster's contribution might just stand alone. In his exquisitely strange, often hilarious chronicle, the writer recounts neuroscientific self-experiments centred on immersion in the sensory maelstrom experienced by iconic British species. He gets down and dirty, chomping earthworms and sleeping in a homemade sett (badger); honing his olfaction and attempting to hunt voles (urban fox); and parachuting (swift). A bold, unsettling try at comprehending the rest of life on Earth.



### Meathooked

Marta Zaraska BASIC (2016)

From char siu to boeuf bourguignon, meat has us hooked, proves journalist Marta Zaraska. Starting 1.5 billion years ago, when one bacterium first engulfed another, she zips through the evolution of human carnivory and examines the enduring pull of animal flesh by way of genetics, developmental biology, chemistry and nutrition. Zaraska negotiates the complexities nimbly, from meat's pivotal part in building our big brains to the 1,000 substances that underpin its cooked odour (3-octen-2-one, for instance, smells of "crushed bugs") and the unpalatable influence of the industry on research.



### Raptor: A Journey Through Birds

James Macdonald Lockhart FOURTH ESTATE (2016)

Fifteen species of raptor — from hen harrier to red kite — form nuclei for the crystalline narratives of this meditation on the British wild, a winner of the 2011 pre-publication Jerwood Award for Non-Fiction. James Macdonald Lockhart grew up knowing birds through photographs shot by his great-grandfather, Seton Gordon. His own understanding of raptor ethology shines. His journey — intercut with passages by Victorian ornithologist William MacGillivray — flings us into skies where a hobby 'concertinas' the air, or a marsh harrier's ruff give it the air of an Elizabethan grandee.



### Eco-Homes: People, Place and Politics

Jenny Pickerill ZED (2016)

We have the science, technology and political will to build eco-housing — as well as to retrofit our not-so-green abodes. What stops us? The greatest hurdle is often cultural adaptation to a new norm, concludes environmental geographer Jenny Pickerill in her cogent sociopolitical work. Focusing on self-build eco-housing, Pickerill looks at more than 30 case studies from a range of countries, along with lessons learned on construction materials, geography and climate, gender, costs and the right policy for rapid rollout. **Barbara Kiser**



Interactions between predators and prey, such as lynxes and hares, can be modelled with biological rules.

# ECOLOGY

# Biology distilled

**Brian J. Enquist** reflects on a blueprint to guide the recovery of life on Earth.

Can biology become as predictive as the physical sciences? And could it guide us in feeding the planet and reversing ecosystem degradation, climate change and species extinction? These are the overarching questions that organize evolutionary biologist Sean Carroll's *The Serengeti Rules* — a compelling read filled with big, bold ideas.

Biology is complicated. A cursory look at a diagram of the biochemical pathways in a cell or interactions between predators and prey in an ecosystem reveals myriad networks. Indeed, following the cascading effect of genes on a phenotype in or between individuals suggests that biological processes are almost unfathomably complex and idiosyncratic. Over the past two decades, those studying such systems have argued that most can be described as a network of interactions. However, this analogy, although powerful, does not necessarily tell us about the resilience of a given system. Nor does it reveal the generalities of the natural world and the regulation of genes, populations and ecosystems.

Carroll tells a richer story. He dispels the idea that biology is too complex to generalize or to become the basis of a predictive science. The book is really about the

unification of biology that has occurred in the shadows of more well known work such as that by Charles Darwin, James Watson and Francis Crick. Through compelling storytelling, key insights of distant, isolated biologists are brought to life.

Carroll finds the common thread in discoveries in anatomy, physiology, gene regulation and cancer research. He does so by way of Nobel-prize-winning molecular biologist Jacques Monod, Janet Rowley, 'mother of chromosome genetics', and ecologists such as Tony Sinclair, who has helped to parse the ecology of the Serengeti region in Tanzania and Kenya.

Carroll distills this body of knowledge into principles. He argues that at all scales of organization, biology is regulated through axioms of interactions in networks — from the number of molecules in our bodies to the numbers and kind of animals and plants in and across ecosystems. He boils all of biology down to six rules of regulation (his



**The Serengeti Rules: The Quest to Discover How Life Works and Why It Matters**  
SEAN B. CARROLL  
Princeton University Press: 2016.

"Serengeti Rules"), which he shows are applicable both to the restoration of ecosystems and to the management of the biosphere.

The same rule may carry different names in different biological contexts. The double-negative logic rule, for instance, enables a given gene product to feed back to slow down its own synthesis. In an ecosystem the same rule, known as top-down regulation, applies when the abundance of a predator (such as lynx) limits the rise in the population of prey (such as snowshoe hares). This is why, in Yellowstone National Park in Wyoming, the reintroduction of wolves has resulted in non-intuitive changes in hydrology and forest cover: wolves prey on elk, which disproportionately feed on streamside willows and tree seedlings. It is also why ecologists can continue to manage the Serengeti, and have been able to 'rebuild' a functioning ecosystem from scratch in Gorongosa National Park, Mozambique.

Carroll argues that the rules regulating human bodily functions — which have improved medical care and driven drug discovery — can be applied to ecosystems, to guide conservation and restoration, and to heal our ailing planet. His *Serengeti Rules* encapsulate the checks and balances that minimize boom-and-bust cycles of species outbreaks and ecosystem imbalances. Ecological systems that are missing key regulatory players, such as predators, can collapse; if they are overtaken by organisms spread by human activities, such as the kudzu vine, a 'cancer-like' growth of that species can result.

Some of Carroll's recommendations are still being debated (see R. D. Grubbs *et al. Sci. Rep.* 6, 20970; 2016). Other pertinent work, such as findings by evolutionary ecologist Daniel Janzen on forest restoration, is already commonly used in conservation. Nonetheless, I suspect that many will find new insights and inspiration here.

Carroll has made a strikingly clear case that ecology is a science on a par with molecular biology and genetics. In many ways, this book is a homage to Charles Elton, who helped to define ecology as the study of species interactions in a 'trophic' network shaped by the environment (see E. Marris *Nature* 459, 327–328; 2009). Building on his vision, Carroll provides a passionate motto for the twenty-first century: "better living through ecology".

Are the *Serengeti Rules* a panacea? No, but Carroll convincingly reveals them to be a sturdy foundation for the future of biology, for human well-being, and for conservation and management. ■

**Brian J. Enquist** is in the department of ecology and evolutionary biology at the University of Arizona in Tucson, and is an external professor at the Santa Fe Institute in New Mexico.  
e-mail: [benquist@email.arizona.edu](mailto:benquist@email.arizona.edu)

➔ **NATURE.COM**  
For more on science in culture see:  
[nature.com/booksandarts](http://nature.com/booksandarts)



# Correspondence

## Japan justifies whaling stance

As Japan's commissioner to the International Whaling Commission (IWC), I disagree that the IWC's review process of scientific whaling is "a waste of time" (A. Brierley and P. Clapham *Nature* **529**, 283; 2016).

The process comprises an independent expert-panel review and a wider review by the IWC Scientific Committee. Research proponents have no say in the expert panel's conclusions. Japan has given due regard to the IWC's criticisms after peer review of its NEWREP-A proposal, which gave the scientific rationale for lethal sampling (see [go.nature.com/wqpxyb](http://go.nature.com/wqpxyb); [go.nature.com/vxjaz6](http://go.nature.com/vxjaz6)).

Brierley and Clapham say that Japan failed to alter its research plans "in any meaningful way" following recommendations by the IWC Scientific Committee that it should explore widely used non-lethal alternatives. In fact, those methods were included in the research plans for evaluation in light of the research objectives. As the International Court of Justice recognized in 2014, certain data cannot be obtained by non-lethal methods (see [go.nature.com/fboxrt](http://go.nature.com/fboxrt)). Japan's new research programme includes both lethal and non-lethal research methods.

The authors' allegation that Japan's whaling is "ostensibly" for research is no basis for proper scientific debate. Japan has made clear that it is always willing to answer questions on its research programme (see [go.nature.com/dut2kx](http://go.nature.com/dut2kx)), and looks forward to constructive scientific discussion at the committee's June meeting. **Joji Morishita** *National Research Institute of Far Seas Fisheries, Shizuoka, Japan.* [jmorishita@affrc.go.jp](mailto:jmorishita@affrc.go.jp)

## NIH push to stop sexual harassment

As the leading US government funder of scientific research, we at the National Institutes of Health

(NIH) are deeply concerned about sexual harassment in science (*Nature* **529**, 255; 2016). With the help of colleagues in government, academia and the private sector, the NIH aims to identify the steps necessary to end this in all NIH-supported research workplaces and scientific meetings.

In September last year, we restated our expectation that organizers of NIH-supported conferences and meetings should assure a safe environment, free of discrimination (see [go.nature.com/zmukk8](http://go.nature.com/zmukk8)).

Over the next few weeks to months, we plan to work with governmental, academic and private-sector colleagues to identify potential steps to translating our expectations into reality. An important first step will be to gather as much data as possible to more fully understand the nature and extent of sexual harassment among scientists. These data should guide us in determining what kinds of policy and procedure are most likely to help. We will also work to determine what levers are already available to influential stakeholders — us as funders, as well as university administrators and departments, journal editors, and organizers and hosts of scientific meetings.

We owe this to our colleagues and the public, who trust in our ability to make the biomedical research enterprise the best that it can be.

**Michael Lauer, Hannah Valantine, Francis S. Collins** *National Institutes of Health, Bethesda, Maryland, USA.* [michael.lauer@nih.gov](mailto:michael.lauer@nih.gov)

## Australians rush to reject primate bill

A bill introduced in the Australian Senate proposes an amendment that would prohibit imports of live non-human primates for research purposes (see *Nature* <http://doi.org/bcqx>; 2016). We call for the Senate

to reject this bill in support of ethically conducted research and preserving the animals' long-term health, for which exchange between international breeding facilities is crucial.

The bill was referred for enquiry to the Senate Legislation Committee for Environment and Communications, which sought public submissions in late 2015 (see [go.nature.com/mjahre](http://go.nature.com/mjahre)). Three days before the first public hearing on 5 February, only 2 out of 56 submissions argued against the amendment.

At that point, we contacted the Australian scientific community — heads of research institutes and those working with non-human primates — and discovered that they were largely unaware of the proposed legislation.

A flurry of written submissions and last-minute personal representations to members of the Senate followed, all calling for rejection of the bill. The international community took up the issue, with many scientific societies and research institutes reacting within 48 hours (see [go.nature.com/oihuxp](http://go.nature.com/oihuxp)).

The Senate committee accepted many submissions after the closing date and is due to submit its report early this month. The episode demonstrates the alacrity with which scientists, typically a reticent group, are prepared to engage with the political process when the issue is perceived as important for the advancement of science.

**Nicholas Price, James Bourne, Marcello Rosa** *Monash University, Melbourne, Australia.* [nicholas.price@monash.edu](mailto:nicholas.price@monash.edu)

## Transparency: issues are not that simple

We find Stephan Lewandowsky and Dorothy Bishop's framing of science governance to be overly simplistic and in need of a firmer evidence base ('Don't let transparency damage science' *Nature* **529**, 459–461; 2016).

The authors' analysis is biased

by its reliance on testimonials from the narrow range of invited experts at last year's Royal Society meeting (see [go.nature.com/zptirs](http://go.nature.com/zptirs)). Complex issues associated with openness and transparency also need to be taken into account (see S. Jasanoff *Law Contemp. Probl.* **69**, 21–45; 2006).

The authors present important topics such as expertise, disciplinary boundaries and communication as simple dichotomies. These divisions overlook extensive nuanced evidence from the social-science literature about who counts as an expert and under which conditions (see, for example, [go.nature.com/xdfrzn](http://go.nature.com/xdfrzn)).

In our view, governance issues around openness and transparency should not be framed only by the research community. The debate must also include representatives from across the broad range of public viewpoints.

**Warren Pearce, Sarah Hartley, Brigitte Nerlich** *University of Nottingham, UK.* [warren.pearce@nottingham.ac.uk](mailto:warren.pearce@nottingham.ac.uk)

## Transparency: an opaque illustration

We question the choice of the padlock and dagger illustration you used to open the discussion by Stephan Lewandowsky and Dorothy Bishop ('Don't let transparency damage science' *Nature* **529**, 459–461; 2016). To us, this falsely implies that the article is about open access to journal publications and, by its association with the title, that open-access publishing presents a threat to science.

The authors send no such scary message, which calls attention to the more general concepts of openness and transparency in providing access to original research data.

**Karen Shashok** *Granada, Spain.* **Remedios Melero** *CSIC Institute of Agrochemistry and Food Technology (IATA), Valencia, Spain.* [kshashok@kshashok.com](mailto:kshashok@kshashok.com)

## ORGANIC CHEMISTRY

## Reactions triggered electrically

Single-molecule experiments have revealed that chemical reactions can be controlled using electric fields — and that the reaction rate is sensitive to both the direction and the strength of the applied field. [SEE LETTER P.88](#)

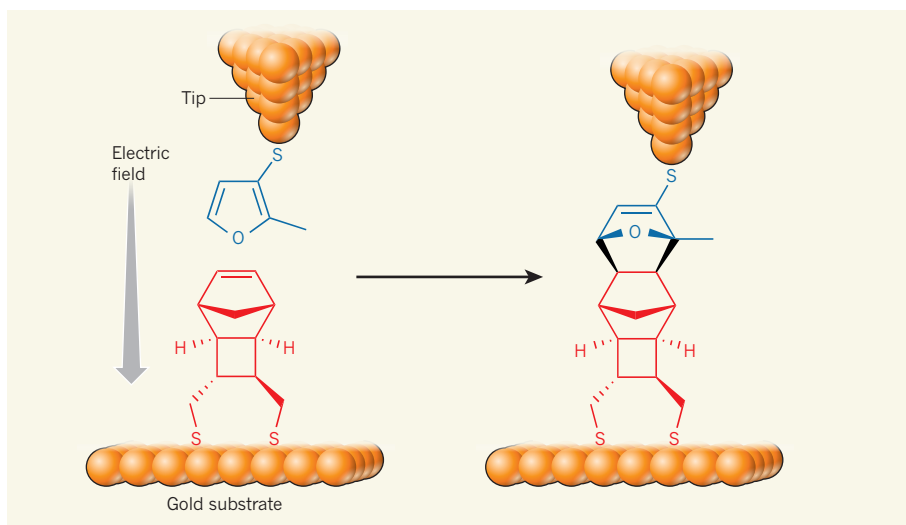
LIMIN XIANG & N. J. TAO

Chemical reactions turn the elements in the periodic table into substances that make up everything we use, ingest and breathe, and into the molecules in our bodies. Developing new and more-efficient ways to control reactions has therefore been a constant quest for chemists. On page 88 of this issue, Aragonès *et al.*<sup>1</sup> describe a new way to control chemical reactions. They report that a classic organic transformation called the Diels–Alder reaction<sup>2</sup> can be accelerated using an electric field, and that the reaction rate sensitively depends not only on the strength of the field, but also on its polarity.

Chemical reactions typically involve the transfer of electrons between atoms and molecules, and rearrangements of the positions of nuclei. To control reactions, one must find a way to promote these processes, and many different driving forces have been used, including heat, light and pressure. It has been predicted theoretically<sup>3</sup> that reactions may also be controlled using an electrostatic field. This is possible because some covalently bonded species can be regarded as combinations of ‘resonance contributors’<sup>4</sup> — molecular structures in which the electrons of chemical bonds can be localized in various ways. A properly oriented electric field can accelerate or decelerate a chemical reaction by stabilizing or destabilizing these contributors.

This mechanism for promoting reactions is distinctly different from that for electrochemical reactions, in which an electric field tunes the energy levels of electrons in molecules such that electron transfer between the electrode and the molecules can occur. It is also different from cases in which an electric field affects a reaction by changing local concentrations of the reactants or catalysts (enzymes), or by bringing charged groups of biological molecules close to each other.

Achieving an electric-field-induced reaction has been a challenge for two main reasons. First, the directions of the electric field and of the molecules must be aligned in a specific way; and second, a detection method must be used that is sensitive enough to measure the enhanced reaction rate. Aragonès *et al.* used a scanning tunnelling microscopy



**Figure 1 | A chemical reaction controlled by an electric field.** Aragonès *et al.*<sup>1</sup> attached a diene molecule (blue) to a sharp gold tip, and a dienophile (red) to a gold substrate. They observed that an applied electric field (arrow) increased the rate of reaction between the molecules, and that the rate increased with the field's strength. An electric field pointing in the opposite direction does not affect the reaction rate (not shown).

break-junction (STM-BJ) technique<sup>5</sup> to create a strong electric field for controlling their Diels–Alder cycloaddition — a reaction that occurs between organic molecules called dienes and dienophiles (Fig. 1). They also used this approach to measure the reaction rate.

The STM-BJ technique was developed to measure the electronic properties of single molecules in a vacuum, in air and in solutions at various temperatures, making it suitable for studying chemical reactions at the molecular level. It involves precisely controlling the separation between a sharp gold tip and a gold substrate. In Aragonès and colleagues' study, the gold tip was modified by attachment of a diene molecule, and the gold surface was modified with a dienophile (Fig. 1). When a voltage was applied between the tip and the substrate, the sharpness of the tip and the small separation between the tip and the substrate helped to generate a strong, directional electric field that aligned the reactants and facilitated a Diels–Alder reaction<sup>6</sup>. Because the STM-BJ technique can detect and analyse a single molecule bridging the tip and substrate, from the current flowing through it, the authors were able to detect the reaction

product at the single-molecule level.

A crucial parameter in the description of a chemical reaction is the reaction rate, which measures how many molecules can be produced in a given time interval. By combining the experimental results with computational approaches<sup>3</sup>, Aragonès *et al.* showed that, by increasing the applied electric field 15-fold, the rate of their reaction increased by up to 5-fold.

Even more interestingly, the researchers observed no effects on the reaction rate in experiments in which the electric field was reversed, demonstrating the importance of the direction of the field. This is because the chemical reaction requires the electrons to flow from the dienophile to the diene, to form a resonance structure in which the dienophile bears a positive charge and the diene carries a negative charge. Only an electric field pointing from the diene to the dienophile can accelerate this process, thereby increasing the reaction rate.

In their experiments, Aragonès and colleagues used the sharp tip and flat substrate as geometrically asymmetric electrodes. This asymmetry might contribute to the reaction rate's dependence on the electric-field



direction. Heat might also be generated when current passes through the molecules during the STM-BJ measurements<sup>7</sup>, which could have an additional effect on the reaction rate. This should be investigated further.

Although many details remain unexplored, the work provides the first experimental evidence that an electric field can control chemical reactions. If this effect can be scaled up for commercially useful reactions on an industrial scale, it could have a huge economic impact. However, the STM-BJ set-up can create a large, directional electric field within only a tiny volume, and would not be suitable for

industrial applications — another technique would need to be developed. Nevertheless, the STM-BJ approach certainly provides a new way to study and control chemical reactions at the single-molecule level, and might provide unprecedented information about reaction mechanisms in the future. ■

**Limin Xiang and N. J. Tao** are at the Center for Biosensors and Bioelectronics, Biodesign Institute, Arizona State University, Tempe, Arizona 85287, USA. **N.J.T.** is also at the State Key Laboratory of Analytical Chemistry for Life Science, Nanjing University, China, and at

the School of Electrical, Computer and Energy Engineering, Arizona State University.  
e-mail: njtao@asu.edu

1. Aragonès, A. C. *et al. Nature* **531**, 88–91 (2016).
2. Nicolaou, K. C., Snyder, S. A., Montagnon, T. & Vassilikogiannakis, G. *Angew. Chem. Int. Edn* **41**, 1668–1698 (2002).
3. Meir, R., Chen, H., Lai, W. & Shaik, S. *ChemPhysChem* **11**, 301–310 (2010).
4. Sini, G., Maitre, P., Hiberty, P. C. & Shaik, S. S. *J. Mol. Struct. THEOCHEM* **229**, 163–188 (1991).
5. Xu, B. & Tao N. J. *Science* **301**, 1221–1223 (2003).
6. Darwish, N., Paddon-Row, M. N. & Gooding, J. J. *Acc. Chem. Res.* **47**, 385–395 (2014).
7. Huang, Z. *et al. Nature Nanotechnol.* **2**, 698–703 (2007).

## EVOLUTION

# Mitochondria in the second act

A large phylogenomics study reveals that the symbiotic event that led to the emergence of organelles known as mitochondria may have occurred later in the evolution of complex cells than was thought. [SEE LETTER P.101](#)

THIJS J. G. ETTEMA

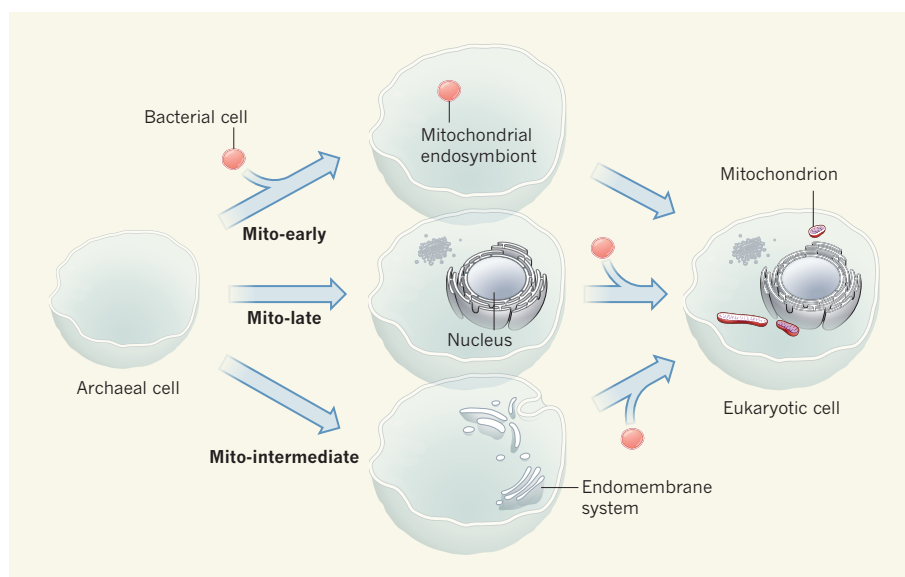
The development of eukaryotic cells, which contain a nucleus and other membrane-bound compartments, is one of the most enigmatic events in the evolution of life on Earth<sup>1</sup>. A crucial moment in this process was the emergence of mitochondria. These organelles are thought to have formed when a bacterial cell began living inside an archaeal host cell — a form of endosymbiosis, a mutually beneficial relationship in which one organism lives inside another. The bacterium is envisaged to have provided its host cell with additional energy<sup>2</sup>, and the interaction eventually resulted in a eukaryotic cell that contained genes of both archaeal and bacterial provenance. In this issue, Pittis and Gabaldón<sup>3</sup> (page 101) provide evidence that the host cell from which eukaryotes evolved was already genetically chimaeric before the mitochondrial symbiosis, suggesting that mitochondria evolved later in eukaryotic evolution than was previously presumed.

Mitochondria generate energy through oxidative-phosphorylation reactions and are therefore sometimes known as the ‘powerhouse’ of eukaryotic cells. Since their discovery more than 100 years ago, the origin of mitochondria has been hotly debated. Currently, an overwhelming amount of evidence indicates that mitochondria are the result of a single endosymbiotic event, and that the mitochondrial progenitor was related to the Alpha-proteobacteria<sup>2,4</sup> (although to which group is still unclear). The archaeal host is thought

to have been related to the Lokiarchaeota, a phylum of archaea that was recognized only last year<sup>5</sup>. Yet there is still uncertainty over when mitochondria emerged during eukaryotic-cell

evolution: was this an early, perhaps even initiating, event (‘mito-early’), or did it occur when the complexity of the eukaryotic cell was already largely established (‘mito-late’)?

According to conventional mito-late models, eukaryotes emerged before the mitochondrial endosymbiont was acquired<sup>7</sup> (Fig. 1). But the popularity of these models has been decreasing with the realization<sup>1</sup> that eukaryotes that do not have mitochondria, and that are thought to have diverged evolutionarily before mitochondria evolved, contain organelles that are degenerate but clearly derived from mitochondria, such as hydrogenosomes and mitosomes. The finding that all known eukaryotes have (or once had) mitochondria resulted in a wave of mito-early hypotheses, in which the interaction between



**Figure 1 | Origin of eukaryotic cells and their mitochondria.** Eukaryotic cells contain membrane-bound compartments, such as a nucleus and mitochondria. The latter are energy-generating organelles that are thought to have formed when a bacterial cell lived in a symbiotic relationship inside an archaeal cell. Models for the origin of eukaryotic cells are usually divided into ‘mito-early’ and ‘mito-late’ scenarios, depending on whether the mitochondrial endosymbiont was acquired early during eukaryotic evolution or when much of the complexity of eukaryotic cells was already established. Pittis and Gabaldón<sup>3</sup> provide evidence for a ‘mito-intermediate’ scenario, in which the cell that hosted the mitochondrial endosymbiont displayed a degree of cellular complexity before the mitochondrial endosymbiosis. But, in contrast to conventional mito-late models, Pittis and Gabaldón’s results do not necessarily imply that the host cell was a fully fledged eukaryote. Hence, their findings are compatible with recent work that supports an archaeal origin for eukaryotes<sup>5,10</sup>.

direction. Heat might also be generated when current passes through the molecules during the STM-BJ measurements<sup>7</sup>, which could have an additional effect on the reaction rate. This should be investigated further.

Although many details remain unexplored, the work provides the first experimental evidence that an electric field can control chemical reactions. If this effect can be scaled up for commercially useful reactions on an industrial scale, it could have a huge economic impact. However, the STM-BJ set-up can create a large, directional electric field within only a tiny volume, and would not be suitable for

industrial applications — another technique would need to be developed. Nevertheless, the STM-BJ approach certainly provides a new way to study and control chemical reactions at the single-molecule level, and might provide unprecedented information about reaction mechanisms in the future. ■

**Limin Xiang and N. J. Tao** are at the Center for Biosensors and Bioelectronics, Biodesign Institute, Arizona State University, Tempe, Arizona 85287, USA. **N.J.T.** is also at the State Key Laboratory of Analytical Chemistry for Life Science, Nanjing University, China, and at

the School of Electrical, Computer and Energy Engineering, Arizona State University.  
e-mail: njtao@asu.edu

1. Aragonès, A. C. *et al. Nature* **531**, 88–91 (2016).
2. Nicolaou, K. C., Snyder, S. A., Montagnon, T. & Vassilikogiannakis, G. *Angew. Chem. Int. Edn* **41**, 1668–1698 (2002).
3. Meir, R., Chen, H., Lai, W. & Shaik, S. *ChemPhysChem* **11**, 301–310 (2010).
4. Sini, G., Maitre, P., Hiberty, P. C. & Shaik, S. S. *J. Mol. Struct. THEOCHEM* **229**, 163–188 (1991).
5. Xu, B. & Tao N. J. *Science* **301**, 1221–1223 (2003).
6. Darwish, N., Paddon-Row, M. N. & Gooding, J. J. *Acc. Chem. Res.* **47**, 385–395 (2014).
7. Huang, Z. *et al. Nature Nanotechnol.* **2**, 698–703 (2007).

## EVOLUTION

# Mitochondria in the second act

A large phylogenomics study reveals that the symbiotic event that led to the emergence of organelles known as mitochondria may have occurred later in the evolution of complex cells than was thought. [SEE LETTER P.101](#)

THIJS J. G. ETTEMA

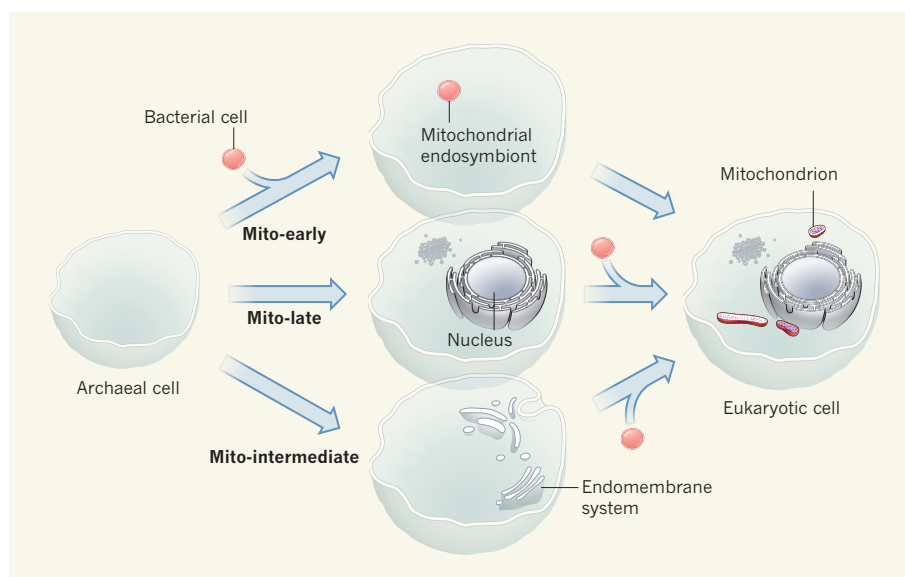
The development of eukaryotic cells, which contain a nucleus and other membrane-bound compartments, is one of the most enigmatic events in the evolution of life on Earth<sup>1</sup>. A crucial moment in this process was the emergence of mitochondria. These organelles are thought to have formed when a bacterial cell began living inside an archaeal host cell — a form of endosymbiosis, a mutually beneficial relationship in which one organism lives inside another. The bacterium is envisaged to have provided its host cell with additional energy<sup>2</sup>, and the interaction eventually resulted in a eukaryotic cell that contained genes of both archaeal and bacterial provenance. In this issue, Pittis and Gabaldón<sup>3</sup> (page 101) provide evidence that the host cell from which eukaryotes evolved was already genetically chimaeric before the mitochondrial symbiosis, suggesting that mitochondria evolved later in eukaryotic evolution than was previously presumed.

Mitochondria generate energy through oxidative-phosphorylation reactions and are therefore sometimes known as the ‘powerhouse’ of eukaryotic cells. Since their discovery more than 100 years ago, the origin of mitochondria has been hotly debated. Currently, an overwhelming amount of evidence indicates that mitochondria are the result of a single endosymbiotic event, and that the mitochondrial progenitor was related to the Alpha-proteobacteria<sup>2,4</sup> (although to which group is still unclear). The archaeal host is thought

to have been related to the Lokiarchaeota, a phylum of archaea that was recognized only last year<sup>5</sup>. Yet there is still uncertainty over when mitochondria emerged during eukaryotic-cell

evolution: was this an early, perhaps even initiating, event (‘mito-early’), or did it occur when the complexity of the eukaryotic cell was already largely established (‘mito-late’)?<sup>6</sup>

According to conventional mito-late models, eukaryotes emerged before the mitochondrial endosymbiont was acquired<sup>7</sup> (Fig. 1). But the popularity of these models has been decreasing with the realization<sup>1</sup> that eukaryotes that do not have mitochondria, and that are thought to have diverged evolutionarily before mitochondria evolved, contain organelles that are degenerate but clearly derived from mitochondria, such as hydrogenosomes and mitosomes. The finding that all known eukaryotes have (or once had) mitochondria resulted in a wave of mito-early hypotheses, in which the interaction between



**Figure 1 | Origin of eukaryotic cells and their mitochondria.** Eukaryotic cells contain membrane-bound compartments, such as a nucleus and mitochondria. The latter are energy-generating organelles that are thought to have formed when a bacterial cell lived in a symbiotic relationship inside an archaeal cell. Models for the origin of eukaryotic cells are usually divided into ‘mito-early’ and ‘mito-late’ scenarios, depending on whether the mitochondrial endosymbiont was acquired early during eukaryotic evolution or when much of the complexity of eukaryotic cells was already established. Pittis and Gabaldón<sup>3</sup> provide evidence for a ‘mito-intermediate’ scenario, in which the cell that hosted the mitochondrial endosymbiont displayed a degree of cellular complexity before the mitochondrial endosymbiosis. But, in contrast to conventional mito-late models, Pittis and Gabaldón’s results do not necessarily imply that the host cell was a fully fledged eukaryote. Hence, their findings are compatible with recent work that supports an archaeal origin for eukaryotes<sup>5,10</sup>.

a primitive host cell and the mitochondrial endosymbiont was the main driving force for eukaryogenesis (Fig. 1).

A syntrophic interaction — in which one species lives off the products of another — is often invoked in these models, and it is thought that the most profound outcome of this interaction was the reallocation of energy production from the host cell's membrane to the mitochondrial membrane. This compartmentalized energy management provided the host with a surplus of energy, which is suggested to have triggered the emergence of the complex cellular features that are characteristic of eukaryotes<sup>8</sup>. The result of this evolutionary journey is proposed to have been the first eukaryotic cell, with a chimaeric genome<sup>9</sup>.

However, although mito-early models have gained much support among evolutionary biologists, the genomic chimaerism in eukaryotes hides a problem: most of the bacterial genes in eukaryotic genomes cannot be traced back to the alleged alphaproteobacterial ancestor of mitochondria. Instead, they seem to originate from various unrelated bacteria. Pittis and Gabaldón aimed to solve this mystery.

By tracing phylogenetic signals of proteins that were present in the last eukaryotic common ancestor (LECA), Pittis and Gabaldón identified different classes of protein according to the timing of their appearance in eukaryotes. In agreement with other findings that imply an archaeal origin for eukaryotes<sup>5,10</sup>, the authors found that the oldest LECA proteins are dominated by archaea-related proteins that are involved in essential cellular functions such as replication, translation and transcription. Furthermore, the most recently acquired LECA proteins are, unsurprisingly, dominated by bacterial proteins, most notably from alphaproteobacteria, that are primarily located in mitochondria and involved in energy generation. Most of these proteins probably originate from the alphaproteobacterial ancestor of mitochondria. Intriguingly, however, Pittis and Gabaldón identified a third class of bacterial LECA protein that they infer was acquired before these mitochondrial proteins. Several of these proteins seem to be located in intracellular membrane systems, such as the endoplasmic reticulum and the Golgi apparatus.

These findings shed light on the relative timing of the origin of mitochondria and the genomic nature of the host cell. First, the results imply that the host cell was already chimaeric before the mitochondrial endosymbiosis. Second, the fact that several of the bacterial proteins that pre-date the mitochondrial endosymbiosis operate in intracellular membrane systems suggests that the host cell already displayed a considerable degree of complexity, which is supportive of a relatively late mitochondrial origin (Fig. 1).

However, Pittis and Gabaldón's results raise a question: what, then, was the origin of the

bacterial genes that pre-date mitochondrial endosymbiosis? Clearly, these genes can no longer be explained by 'inherited chimaerism' of the mitochondrial endosymbiont<sup>11</sup>. The authors suggest that the genes may have been acquired through previous (endo)symbiotic interactions with different bacterial partners or by serial waves of horizontal gene transfer to the host genome.

Although this question remains open, clues to an answer might come from genomes of the Lokiarchaeota phylum, members of which share a common ancestry with eukaryotes. Analysis of a lokiarchaeal genome indicated that nearly 30% of its genes display greater similarity to bacterial than to archaeal genes<sup>5</sup>. Despite being extensively reshaped by evolutionary processes, some of the bacterial genes in the archaeal ancestor of eukaryotes could have ended up in the genomes of present-day eukaryotes. Future exploration of new lineages of archaea and eukaryotic microorganisms will provide yet more insight into the origin and early evolution of eukaryotes,

including the key role of mitochondria in this process. ■

**Thijs J. G. Ettema** is in the Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-75123 Uppsala, Sweden.  
e-mail: [thijs.ettema@icm.uu.se](mailto:thijs.ettema@icm.uu.se)

1. Embley, T. M. & Martin, W. *Nature* **440**, 623–630 (2006).
2. Gray, M. W., Burger, G. & Lang, B. F. *Science* **283**, 1476–1481 (1999).
3. Pittis, A. A. & Gabaldón, T. *Nature* **531**, 101–104 (2016).
4. Rochette, N. C., Brochier-Armanet, C. & Gouy, M. *Mol. Biol. Evol.* **31**, 832–845 (2014).
5. Spang, A. et al. *Nature* **521**, 173–179 (2015).
6. Poole, A. M. & Gribaldo, S. *Cold Spring Harb. Perspect. Biol.* **6**, a015990 (2014).
7. Cavalier-Smith, T. *Nature* **339**, 100–101 (1989).
8. Lane, N. & Martin, W. *Nature* **467**, 929–934 (2010).
9. Koonin, E. V. *Genome Biol.* **11**, 209 (2010).
10. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. *Nature* **504**, 231–236 (2013).
11. Ku, C. et al. *Proc. Natl Acad. Sci. USA* **112**, 10139–10146 (2015).

This article was published online on 3 February 2016.

#### GRAVITATIONAL WAVES

## Dawn of a new astronomy

The discovery of gravitational waves from a merging black-hole system opens a window on the Universe that promises to test gravity at its strongest, and to reveal many surprises about black holes and other astrophysical systems.

M. COLEMAN MILLER

Shortly after Albert Einstein delivered his theory of gravity — general relativity — to the world in 1915, he discovered that binary stars and other sources should generate gravitational waves<sup>1,2</sup>. Unfortunately, he also found that any imaginable source would produce gravitational waves so weak that detection was inconceivable using the technology of the day. But this inconceivable detection has now been reported by Abbott *et al.* (the LIGO Scientific Collaboration and the Virgo Collaboration) in *Physical Review Letters*<sup>3</sup>.

The authors describe the detection of the signal GW150914 from gravitational waves generated by the merger of two black holes (Fig. 1). These waves were detected from the temporary, tiny changes that they induced in the lengths of the two detectors of the US-based Advanced Laser Interferometer Gravitational-Wave Observatory (Advanced LIGO). The consequences of this detection are difficult to overstate, as is its promise for future advances and discoveries. Before this discovery, astronomers had only three types of messenger from space

beyond our Solar System: photons, neutrinos and high-energy cosmic rays. Gravitational waves can now be added to this short list.

Moreover, some of the most violent events in the Universe can be seen only in gravitational waves. Consider, for example, the inspiral and merger of two black holes, such as the one that caused GW150914. During the last moments of coalescence, the energy emitted in gravitational waves was tens of times larger than the energy emitted in those moments by all the stars in the Universe combined. But such an event is expected to be undetectable using any of the other messengers. Opening the gravitational-wave window will thus reveal events that had previously only been hypothesized.

This window will also enable tests of general relativity in realms that were previously inaccessible. We can get an idea of the domain that can now be explored by considering a dimensionless quantity,  $GM/Rc^2$ , that measures the importance of gravity for an object of mass  $M$  and radius  $R$  ( $G$  is Newton's gravitational constant and  $c$  is the speed of light). For Earth,  $GM/Rc^2$  at the surface is approximately  $7 \times 10^{-10}$ . For a star such as the Sun,  $GM/Rc^2$



a primitive host cell and the mitochondrial endosymbiont was the main driving force for eukaryogenesis (Fig. 1).

A syntrophic interaction — in which one species lives off the products of another — is often invoked in these models, and it is thought that the most profound outcome of this interaction was the reallocation of energy production from the host cell's membrane to the mitochondrial membrane. This compartmentalized energy management provided the host with a surplus of energy, which is suggested to have triggered the emergence of the complex cellular features that are characteristic of eukaryotes<sup>8</sup>. The result of this evolutionary journey is proposed to have been the first eukaryotic cell, with a chimaeric genome<sup>9</sup>.

However, although mito-early models have gained much support among evolutionary biologists, the genomic chimaerism in eukaryotes hides a problem: most of the bacterial genes in eukaryotic genomes cannot be traced back to the alleged alphaproteobacterial ancestor of mitochondria. Instead, they seem to originate from various unrelated bacteria. Pittis and Gabaldón aimed to solve this mystery.

By tracing phylogenetic signals of proteins that were present in the last eukaryotic common ancestor (LECA), Pittis and Gabaldón identified different classes of protein according to the timing of their appearance in eukaryotes. In agreement with other findings that imply an archaeal origin for eukaryotes<sup>5,10</sup>, the authors found that the oldest LECA proteins are dominated by archaea-related proteins that are involved in essential cellular functions such as replication, translation and transcription. Furthermore, the most recently acquired LECA proteins are, unsurprisingly, dominated by bacterial proteins, most notably from alphaproteobacteria, that are primarily located in mitochondria and involved in energy generation. Most of these proteins probably originate from the alphaproteobacterial ancestor of mitochondria. Intriguingly, however, Pittis and Gabaldón identified a third class of bacterial LECA protein that they infer was acquired before these mitochondrial proteins. Several of these proteins seem to be located in intracellular membrane systems, such as the endoplasmic reticulum and the Golgi apparatus.

These findings shed light on the relative timing of the origin of mitochondria and the genomic nature of the host cell. First, the results imply that the host cell was already chimaeric before the mitochondrial endosymbiosis. Second, the fact that several of the bacterial proteins that pre-date the mitochondrial endosymbiosis operate in intracellular membrane systems suggests that the host cell already displayed a considerable degree of complexity, which is supportive of a relatively late mitochondrial origin (Fig. 1).

However, Pittis and Gabaldón's results raise a question: what, then, was the origin of the

bacterial genes that pre-date mitochondrial endosymbiosis? Clearly, these genes can no longer be explained by 'inherited chimaerism' of the mitochondrial endosymbiont<sup>11</sup>. The authors suggest that the genes may have been acquired through previous (endo)symbiotic interactions with different bacterial partners or by serial waves of horizontal gene transfer to the host genome.

Although this question remains open, clues to an answer might come from genomes of the Lokiarchaeota phylum, members of which share a common ancestry with eukaryotes. Analysis of a lokiarchaeal genome indicated that nearly 30% of its genes display greater similarity to bacterial than to archaeal genes<sup>5</sup>. Despite being extensively reshaped by evolutionary processes, some of the bacterial genes in the archaeal ancestor of eukaryotes could have ended up in the genomes of present-day eukaryotes. Future exploration of new lineages of archaea and eukaryotic microorganisms will provide yet more insight into the origin and early evolution of eukaryotes,

including the key role of mitochondria in this process. ■

**Thijs J. G. Ettema** is in the Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-75123 Uppsala, Sweden.  
e-mail: thijs.ettema@icm.uu.se

1. Embley, T. M. & Martin, W. *Nature* **440**, 623–630 (2006).
2. Gray, M. W., Burger, G. & Lang, B. F. *Science* **283**, 1476–1481 (1999).
3. Pittis, A. A. & Gabaldón, T. *Nature* **531**, 101–104 (2016).
4. Rochette, N. C., Brochier-Armanet, C. & Gouy, M. *Mol. Biol. Evol.* **31**, 832–845 (2014).
5. Spang, A. et al. *Nature* **521**, 173–179 (2015).
6. Poole, A. M. & Gribaldo, S. *Cold Spring Harb. Perspect. Biol.* **6**, a015990 (2014).
7. Cavalier-Smith, T. *Nature* **339**, 100–101 (1989).
8. Lane, N. & Martin, W. *Nature* **467**, 929–934 (2010).
9. Koonin, E. V. *Genome Biol.* **11**, 209 (2010).
10. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. *Nature* **504**, 231–236 (2013).
11. Ku, C. et al. *Proc. Natl Acad. Sci. USA* **112**, 10139–10146 (2015).

This article was published online on 3 February 2016.

#### GRAVITATIONAL WAVES

## Dawn of a new astronomy

The discovery of gravitational waves from a merging black-hole system opens a window on the Universe that promises to test gravity at its strongest, and to reveal many surprises about black holes and other astrophysical systems.

M. COLEMAN MILLER

Shortly after Albert Einstein delivered his theory of gravity — general relativity — to the world in 1915, he discovered that binary stars and other sources should generate gravitational waves<sup>1,2</sup>. Unfortunately, he also found that any imaginable source would produce gravitational waves so weak that detection was inconceivable using the technology of the day. But this inconceivable detection has now been reported by Abbott *et al.* (the LIGO Scientific Collaboration and the Virgo Collaboration) in *Physical Review Letters*<sup>3</sup>.

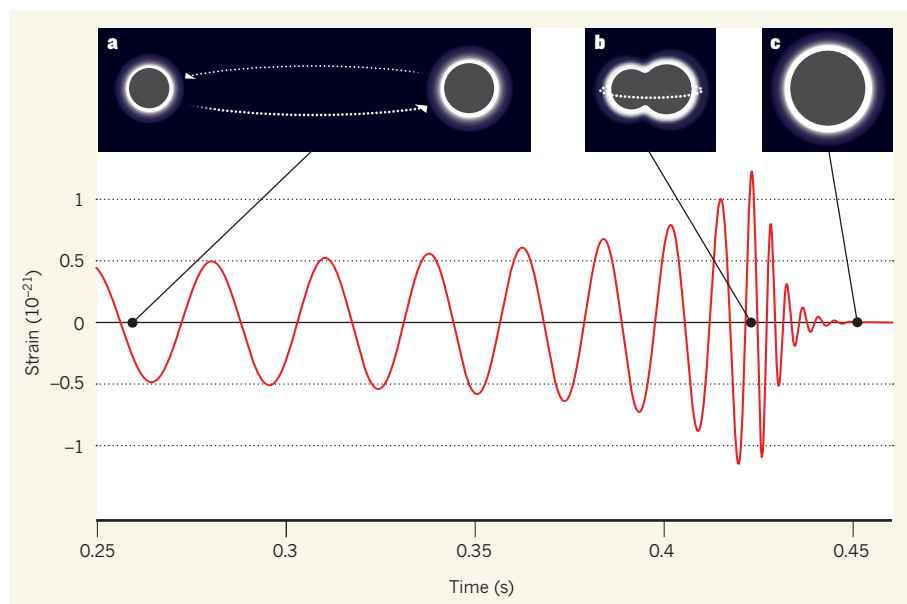
The authors describe the detection of the signal GW150914 from gravitational waves generated by the merger of two black holes (Fig. 1). These waves were detected from the temporary, tiny changes that they induced in the lengths of the two detectors of the US-based Advanced Laser Interferometer Gravitational-Wave Observatory (Advanced LIGO). The consequences of this detection are difficult to overstate, as is its promise for future advances and discoveries. Before this discovery, astronomers had only three types of messenger from space

beyond our Solar System: photons, neutrinos and high-energy cosmic rays. Gravitational waves can now be added to this short list.

Moreover, some of the most violent events in the Universe can be seen only in gravitational waves. Consider, for example, the inspiral and merger of two black holes, such as the one that caused GW150914. During the last moments of coalescence, the energy emitted in gravitational waves was tens of times larger than the energy emitted in those moments by all the stars in the Universe combined. But such an event is expected to be undetectable using any of the other messengers. Opening the gravitational-wave window will thus reveal events that had previously only been hypothesized.

This window will also enable tests of general relativity in realms that were previously inaccessible. We can get an idea of the domain that can now be explored by considering a dimensionless quantity,  $GM/Rc^2$ , that measures the importance of gravity for an object of mass  $M$  and radius  $R$  ( $G$  is Newton's gravitational constant and  $c$  is the speed of light). For Earth,  $GM/Rc^2$  at the surface is approximately  $7 \times 10^{-10}$ . For a star such as the Sun,  $GM/Rc^2$





**Figure 1 | A gravitational wave from merging black holes.** Abbott *et al.*<sup>3</sup> report the detection of a gravitational wave, which they attribute to the coalescence of two black holes. **a**, The wave was first detected at approximately 35 Hz, as it reached the sensitivity range of Advanced LIGO (the detecting observatory). At this point, the black holes were spiralling in towards each other. The depicted radii are proportional to the black-holes' masses. **b**, The wave frequency increased as the black holes coalesced — at the point of merger, the black-hole horizons overlapped, but had not settled down to their final state. **c**, The wave dissipated as the merged black hole attained its final, simple configuration. The wave depicted here is based on observational data, but has been smoothed and fitted to a numerical model based on general relativity; strain represents the fractional changes in distance that are produced by the waves. (Adapted from ref. 1.)

is still only about  $2 \times 10^{-6}$ . Previous tests of general relativity have therefore been restricted to systems that have weak gravity.

But at the event horizon of a black hole (the boundary beyond which nothing can escape the hole's gravitational field),  $GM/Rc^2$  is roughly 1, many orders of magnitude larger than for planets and stars. Gravity can thus be tested directly at its greatest strength for the first time, by analysing GW150914 and any other signals detected for similar mergers in the future. General relativity has passed the tests set by GW150914 with flying colours<sup>4</sup>. This signal has also provided the most direct confirmation yet of the existence of event horizons, which are unique to black holes.

The discovery of GW150914 has profound consequences for astronomy. Previously known black holes that formed from a single star have quite a restricted mass range: the highest mass that was definitively established was found to be only about 15 times the mass of the Sun<sup>5</sup>. Analysis of GW150914 has doubled this mass record at a stroke (the merging black holes had masses 29 and 36 times that of the Sun<sup>6</sup>), and then doubled it again (the final merged black hole is inferred to have a mass 62 times that of the Sun<sup>6</sup>). The spins of black holes are notoriously difficult to measure, but Abbott *et al.* were able to infer the spin of the final black hole from their data: the 160-kilometre-radius black hole spins completely around 100 times per second, which is roughly 70% of the maximum possible rate for a black hole of this mass.

None of this could have happened without spectacular developments in instrumentation. Gravitational waves distort space and time only slightly at our distance from any likely sources. The distortion is characterized by a dimensionless quantity called strain, which is the fractional change in distances produced by the waves. Even for a fairly strong event such as the black-hole merger, the change is tiny: the authors find a maximum value of just  $10^{-21}$ . This means that the 4-kilometre-long, L-shaped arms of Advanced LIGO change in length by about 1/200 of the radius of a proton. Such changes can nonetheless be seen because of the exquisite precision of the optics of Advanced LIGO, the delicacy of its suspension and the power of its lasers, which all result from years of development — the LIGO detector has improved in all respects by orders of magnitude since its first conception more than 40 years ago.

Even more encouragingly, further major improvements are just around the corner. During its next run in 2016, Advanced LIGO will be able to observe about three times the volume of space that it could in 2015, and in the next year or two the Advanced Virgo detector in Italy will join the search for gravitational waves. A few years later, the Japanese Kamioka Gravitational Wave Detector will come online, and it is hoped that LIGO-India will join the hunt before 2025. This international network will also benefit from technological developments in light manipulation such as those at the GEO600 detector in Germany. The



## 50 Years Ago

The dilatory attitude of the Committee on Libraries set up by the University Grants Committee and the Ministry of Education, and the naïve remark in the Robbins Report that “a library adequate to scholarly research is as essential to the efficient running of a university as an adequate range of computers”, fully justifies the anxiety expressed recently by the Master of Sidney Sussex College, Cambridge ... Universities, as Dr. D. Thomson pointed out, have managed for centuries without computers, but never without libraries.

From *Nature* 5 March 1966

## 100 Years Ago

We notice in *La Geographie* for November, 1915, that the hydrographic department of the French Admiralty have replaced the German names in Kerguelen by names of French origin. It must be very galling to the French to see an abundance of German names scattered over the chart of their Antarctic island, especially as German explorers were never very sparing in their naming ... however, the practice of changing established names is a dangerous one if carried far, and it is to be hoped ... this principle will not be applied indiscriminately, for confusion would certainly be the result.

### ALSO:

Those who are inclined to doubt whether museums play any useful part in war-time should read the account of what is being done in the Leicester Museum, by means of an Infant Welfare Exhibition, to combat the appalling mortality among infants ... This mortality, which is largely preventable, is brought out with startling vividness by means of a series of wooden columns, that for infants up to twelve months old standing no fewer than 11 ft. high, while that for the death-rate between the ages from five to twenty is but 2¾ of an inch high.

From *Nature* 2 March 1916

resulting volume of space that can be explored will be tens of times greater than could be seen during the GW150914 detection, and will allow the direction of future events to be determined much more accurately than was possible for GW150914.

Surprises undoubtedly await, particularly given that the ability to detect gravitational waves at new frequency ranges is being developed in facilities such as the space-based Evolved Laser Interferometer Space Antenna and the ground-based International Pulsar Timing Array, and for various experiments that are studying the polarization of the cosmic

microwave background (the oldest light in the Universe). Just as when Galileo turned his telescope to the heavens for the first time, everything will be new. It is truly a privilege to be present at the dawn of gravitational-wave astronomy. ■

**M. Coleman Miller** is in the Department of Astronomy and Joint Space-Science Institute, University of Maryland, College Park, Maryland 20742-2421, USA.  
e-mail: miller@astro.umd.edu

1. Einstein, A. *Sitzungsber. K. Preuss. Akad. Wiss.* **1**, 688 (1916).

- Einstein, A. *Sitzungsber. K. Preuss. Akad. Wiss.* **1**, 154 (1918).
- Abbott, B. P. *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) *Phys. Rev. Lett.* **116**, 061102 (2016).
- The LIGO Scientific Collaboration, the Virgo Collaboration. Preprint at <http://arxiv.org/abs/1602.03841> (2016).
- Kreidberg, L., Bailyn, C. D., Farr, W. M. & Kalogera, V. *Astrophys. J.* **757**, 36 (2012).
- The LIGO Scientific Collaboration, the Virgo Collaboration. Preprint at <http://arxiv.org/abs/1602.03840>.

The author declares competing interests. See [go.nature.com/wfoqkb](http://go.nature.com/wfoqkb) for details.

This article was published online on 24 February 2016.

## STEM CELLS

# Dietary fat promotes intestinal dysregulation

In mice, a high-fat diet has now been found to induce intestinal progenitor cells to adopt a more stem-cell-like fate, altering the size of the gut and increasing tumour incidence. [SEE ARTICLE P.53](#)

CHI LUO & PERE PUIGSERVER

Excess food intake causes obesity and is linked to many life-threatening diseases, including cardiovascular disease and cancer<sup>1</sup>. Ingested dietary components cause a well-understood physiological response that readjusts metabolism and restores energy balance, and that is controlled by nutrients, hormones and systemic factors<sup>2</sup>. But little is known about how dietary components might affect stem-cell biology to alter tissue function or tumour development<sup>3</sup>. On page 53 of this issue, Beyaz *et al.*<sup>4</sup> show that an increase in dietary fat directly promotes the proliferation of intestinal stem cells (ISCs) and the progenitor cells to which they give rise, perhaps resulting in more 'seeds' that can develop into cancer.

Tissue homeostasis in the gut requires that resident ISCs maintain a dynamic balance between self-renewal, which expands the stem-cell pool, and differentiation into daughters — intestinal progenitors that eventually give rise to all the mature lineages of the intestinal epithelium (the intestinal lining). Nutrients cause changes in a variety of circulating factors that influence adult stem-cell biology, affecting this balance and so altering tissue remodelling and regeneration<sup>2</sup>. In addition, ISCs are in close contact with dietary constituents undergoing digestion, and so are directly regulated by molecular components of ingested food.

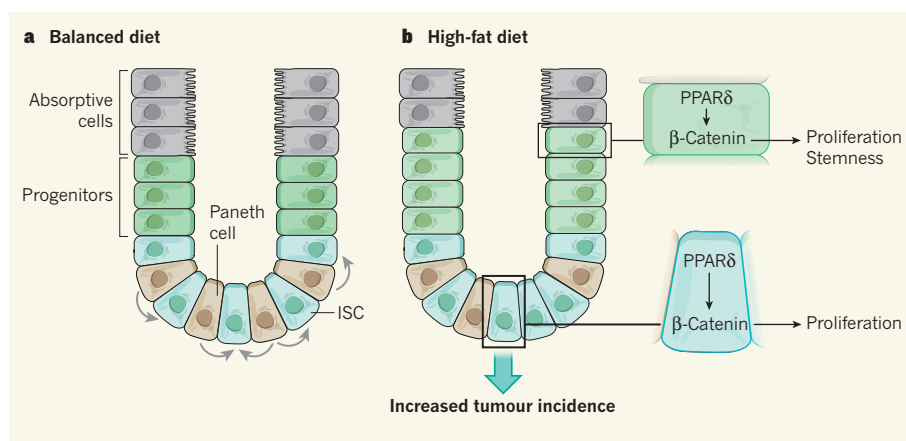
Beyaz *et al.* report that a high-fat diet (HFD) elevates the number and proliferation rate of both ISCs and progenitors in mice. This

alteration led to elongation and regeneration of pits in the epithelium called crypts, in which these cell types are located. The HFD augmented the ability of intestinal crypts to give rise to mini-gut-like structures called organoids when cultured *in vitro* — an approach that is widely used to assay ISC activity. Progenitors from HFD mice could also form organoids, suggesting that they become more stem-cell-like under HFD conditions (Fig. 1). Rather than being a result of obesity per se, these changes in ISC and progenitor behaviour

were caused by certain fatty acids in the HFD.

Despite the changes in ISC and progenitor function, Beyaz and colleagues found that the overall length and weight of the intestine were reduced in mice fed an HFD, compared with control animals. And, explaining this finding, the number of certain mature cell types — absorptive cells and Paneth cells, which defend against harmful bacteria in the gut — was lower in HFD mice. Together, these findings imply that an undifferentiated ISC pool is maintained despite the expanded numbers and increased regenerative potential of the ISCs.

Stem cells normally reside in a specialized environment called a niche, in which communication with neighbouring cells ensures their precise regulation. Paneth cells are an essential part of the niche<sup>5</sup>, and are interspersed throughout it. A previous study<sup>6</sup> from the same group revealed that calorie restriction increases the number of niche Paneth cells, promoting ISC self-renewal and subsequent intestinal regeneration. By contrast, in the current study, the authors demonstrated that an unrestricted HFD uncoupled the ISCs from their niche, allowing them to adjust to the



**Figure 1 | Dietary fats remodel the intestine.** Intestinal stem cells (ISCs) and their daughters, intestinal progenitors, reside at the bottom of structures called crypts, and give rise to all the cell types of the mature intestine, including Paneth cells and absorptive cells. When mice are fed a balanced diet, signals released by Paneth cells (grey arrows) promote ISC regeneration. Beyaz *et al.*<sup>4</sup> report that some of the fatty acids in a high-fat diet (HFD) activate a signalling cascade that involves the nuclear receptor protein PPAR- $\delta$  and the protein  $\beta$ -catenin. The cascade increases ISC and progenitor proliferation, makes progenitors more stem-cell-like and enables ISCs to grow in the absence of signals from Paneth cells. This HFD-mediated expansion of the stem and progenitor pool promotes tumour formation.

resulting volume of space that can be explored will be tens of times greater than could be seen during the GW150914 detection, and will allow the direction of future events to be determined much more accurately than was possible for GW150914.

Surprises undoubtedly await, particularly given that the ability to detect gravitational waves at new frequency ranges is being developed in facilities such as the space-based Evolved Laser Interferometer Space Antenna and the ground-based International Pulsar Timing Array, and for various experiments that are studying the polarization of the cosmic

microwave background (the oldest light in the Universe). Just as when Galileo turned his telescope to the heavens for the first time, everything will be new. It is truly a privilege to be present at the dawn of gravitational-wave astronomy. ■

**M. Coleman Miller** is in the Department of Astronomy and Joint Space-Science Institute, University of Maryland, College Park, Maryland 20742-2421, USA.  
e-mail: miller@astro.umd.edu

1. Einstein, A. *Sitzungsber. K. Preuss. Akad. Wiss.* **1**, 688 (1916).

- Einstein, A. *Sitzungsber. K. Preuss. Akad. Wiss.* **1**, 154 (1918).
- Abbott, B. P. *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) *Phys. Rev. Lett.* **116**, 061102 (2016).
- The LIGO Scientific Collaboration, the Virgo Collaboration. Preprint at <http://arxiv.org/abs/1602.03841> (2016).
- Kreidberg, L., Bailyn, C. D., Farr, W. M. & Kalogera, V. *Astrophys. J.* **757**, 36 (2012).
- The LIGO Scientific Collaboration, the Virgo Collaboration. Preprint at <http://arxiv.org/abs/1602.03840>.

The author declares competing interests. See [go.nature.com/wfoqkb](http://go.nature.com/wfoqkb) for details.

This article was published online on 24 February 2016.

## STEM CELLS

# Dietary fat promotes intestinal dysregulation

In mice, a high-fat diet has now been found to induce intestinal progenitor cells to adopt a more stem-cell-like fate, altering the size of the gut and increasing tumour incidence. [SEE ARTICLE P.53](#)

CHI LUO & PERE PUIGSERVER

Excess food intake causes obesity and is linked to many life-threatening diseases, including cardiovascular disease and cancer<sup>1</sup>. Ingested dietary components cause a well-understood physiological response that readjusts metabolism and restores energy balance, and that is controlled by nutrients, hormones and systemic factors<sup>2</sup>. But little is known about how dietary components might affect stem-cell biology to alter tissue function or tumour development<sup>3</sup>. On page 53 of this issue, Beyaz *et al.*<sup>4</sup> show that an increase in dietary fat directly promotes the proliferation of intestinal stem cells (ISCs) and the progenitor cells to which they give rise, perhaps resulting in more 'seeds' that can develop into cancer.

Tissue homeostasis in the gut requires that resident ISCs maintain a dynamic balance between self-renewal, which expands the stem-cell pool, and differentiation into daughters — intestinal progenitors that eventually give rise to all the mature lineages of the intestinal epithelium (the intestinal lining). Nutrients cause changes in a variety of circulating factors that influence adult stem-cell biology, affecting this balance and so altering tissue remodelling and regeneration<sup>2</sup>. In addition, ISCs are in close contact with dietary constituents undergoing digestion, and so are directly regulated by molecular components of ingested food.

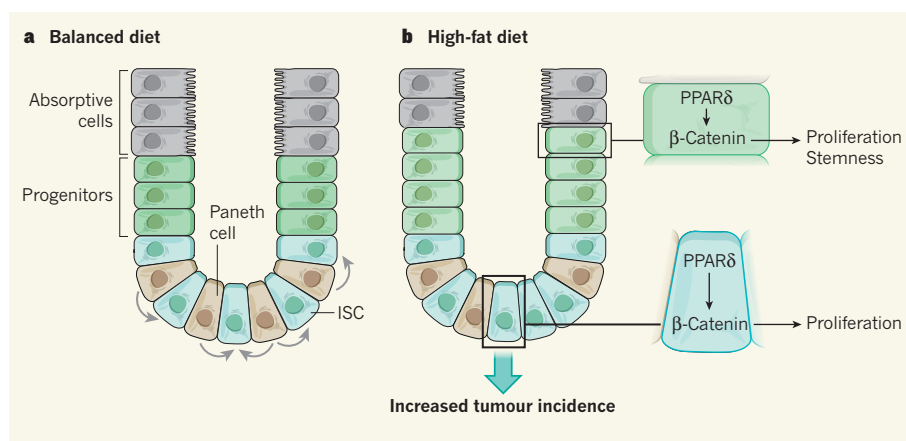
Beyaz *et al.* report that a high-fat diet (HFD) elevates the number and proliferation rate of both ISCs and progenitors in mice. This

alteration led to elongation and regeneration of pits in the epithelium called crypts, in which these cell types are located. The HFD augmented the ability of intestinal crypts to give rise to mini-gut-like structures called organoids when cultured *in vitro* — an approach that is widely used to assay ISC activity. Progenitors from HFD mice could also form organoids, suggesting that they become more stem-cell-like under HFD conditions (Fig. 1). Rather than being a result of obesity per se, these changes in ISC and progenitor behaviour

were caused by certain fatty acids in the HFD.

Despite the changes in ISC and progenitor function, Beyaz and colleagues found that the overall length and weight of the intestine were reduced in mice fed an HFD, compared with control animals. And, explaining this finding, the number of certain mature cell types — absorptive cells and Paneth cells, which defend against harmful bacteria in the gut — was lower in HFD mice. Together, these findings imply that an undifferentiated ISC pool is maintained despite the expanded numbers and increased regenerative potential of the ISCs.

Stem cells normally reside in a specialized environment called a niche, in which communication with neighbouring cells ensures their precise regulation. Paneth cells are an essential part of the niche<sup>5</sup>, and are interspersed throughout it. A previous study<sup>6</sup> from the same group revealed that calorie restriction increases the number of niche Paneth cells, promoting ISC self-renewal and subsequent intestinal regeneration. By contrast, in the current study, the authors demonstrated that an unrestricted HFD uncoupled the ISCs from their niche, allowing them to adjust to the



**Figure 1 | Dietary fats remodel the intestine.** Intestinal stem cells (ISCs) and their daughters, intestinal progenitors, reside at the bottom of structures called crypts, and give rise to all the cell types of the mature intestine, including Paneth cells and absorptive cells. When mice are fed a balanced diet, signals released by Paneth cells (grey arrows) promote ISC regeneration. Beyaz *et al.*<sup>4</sup> report that some of the fatty acids in a high-fat diet (HFD) activate a signalling cascade that involves the nuclear receptor protein PPAR- $\delta$  and the protein  $\beta$ -catenin. The cascade increases ISC and progenitor proliferation, makes progenitors more stem-cell-like and enables ISCs to grow in the absence of signals from Paneth cells. This HFD-mediated expansion of the stem and progenitor pool promotes tumour formation.



decrease in Paneth-cell numbers. For instance, several signalling proteins (such as Jag1 and Jag2) that are normally produced by Paneth cells are upregulated in HFD-derived ISCs, sustaining niche-independent growth.

The incidence of human colorectal cancer correlates with diet-induced obesity<sup>7</sup>. Furthermore, adult stem cells are speculated to be the origin of some cancers<sup>8</sup>. Beyaz *et al.* showed that the increased pool of ISCs and ISC-like progenitors induced by an HFD predisposed mice to intestinal tumours. By contrast, calorie restriction — which also increases ISC numbers<sup>6</sup> — is associated with reduced tumour initiation<sup>1</sup>. The mechanistic differences underlying altered stem-cell function in each condition may partially explain this discrepancy. Calorie restriction is associated with increased interactions between the niche and ISCs, whereas HFD-associated, niche-independent growth allows stem cells to escape physiological regulation. How the molecular pathways modulated by these two dietary regimens intersect and communicate in ISCs remains to be investigated, and might identify putative therapeutic targets.

Gene-expression profiles often provide clues to the state of a cell. Nuclear receptors such as PPAR and LXR proteins sense nutrients and regulate gene expression, providing a connection between diet-induced metabolic changes and these profiles<sup>9</sup>. Beyaz and colleagues analysed gene expression and found that genetic targets of one PPAR, PPAR- $\delta$ , were upregulated in ISCs in HFD mice compared to ISCs in control mice. PPAR- $\delta$  is linked to a signalling cascade called the Wnt- $\beta$ -catenin pathway<sup>10</sup>, which is involved in the development of intestinal tumours. In the current study, genetic and pharmacological experiments revealed that activation of PPAR- $\delta$  and Wnt- $\beta$ -catenin signalling mediates, at least in part, the effects of dietary fats on ISC and progenitor function and intestinal tumour formation.

This work provides a plausible cellular and molecular explanation for how an excess of dietary fat remodels the intestine. However, questions remain about how the basic mechanisms of action of dietary fat affect systemic energy metabolism and other gastrointestinal diseases apart from cancer. For instance, it is known that diet can influence immune and metabolic activities by directly modulating the diversity and functions of the gut microbiota (the population of microorganisms that inhabit the gut), but it remains unclear whether and how such changes in the microbiota integrate with the PPAR- $\delta$  pathway. Because the microbiota differs between individuals, the interplay between microbiota and ISCs under HFD conditions might modulate an individual's tumour risk.

It would be interesting to investigate the contribution of ISCs to gut inflammatory disorders such as Crohn's disease because, as with cancer, an HFD accelerates the progress

of these disorders independently of obesity<sup>11</sup>. Furthermore, it is not known whether an HFD affects the gut's neuroendocrine system (the hormone-releasing cells that receive input from neurons), perhaps through its effects on ISCs and progenitors, to contribute to the metabolic alterations associated with obesity, type 2 diabetes or cardiovascular diseases.

The current study does not address whether the effects of an HFD on gut architecture are reversible. Moreover, it is unclear how changes in dietary regimens affect ISC function. Finally, further investigation will be needed to determine whether dietary or pharmacological interventions that target ISCs could maintain healthy intestinal function and reduce the incidence of tumours or other HFD-associated human diseases. Such research, building on the foundation provided by the current study, will be important for defining future steps in personalized human nutrition and health. ■

## SPACE SCIENCE

# Cosmic rays beyond the knees

**The development of a radio technique for detecting cosmic rays casts fresh light on the origins of some of these accelerated particles, and suggests that they might have travelled much farther than was previously thought. [SEE LETTER P.70](#)**

ANDREW M. TAYLOR

A technique for measuring the composition of cosmic rays in the energy range  $10^{17}$  to  $10^{17.5}$  electronvolts — a range thought to mark the transition point from Galactic to extragalactic cosmic rays — is reported by Buitink *et al.*<sup>1</sup> on page 70 of this issue. The findings have implications for our understanding of the sources of these mysterious rays.

Energetic protons and atomic nuclei arriving at Earth are classified as cosmic rays. The spread in energy of these particles is impressive, covering ten orders of magnitude (Fig. 1). The most energetic cosmic rays have energies more than 10 million times those achievable for protons at the Large Hadron Collider, the world's most powerful particle accelerator at CERN in Geneva, Switzerland. This begs the question of what cosmic sources can accelerate particles to such high energies.

Clues about the origin of cosmic rays come from both their composition and their energy spectra. Unfortunately, these particles arrive at Earth rather infrequently — only one particle with an energy greater than  $10^{17}$  eV arrives each day in every 10,000 square metres. This scarcity makes it difficult to build detectors

**Chi Luo and Pere Puigserver** are in the Department of Cell Biology, Harvard Medical School, and in the Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. e-mail: pere\_puigserver@dfci.harvard.edu

1. Nihaylova, M. M., Sabatini, D. M. & Yilmaz, O. H. *Cell Stem Cell* **14**, 292–305 (2014).
2. Nakada, D., Levi, B. P. & Morrison, S. J. *Neuron* **70**, 703–718 (2011).
3. Shyh-Chang, N., Daley, G. Q. & Cantley, L. C. *Development* **140**, 2535–2547 (2013).
4. Beyaz, S. *et al.* *Nature* **531**, 53–58 (2016).
5. Sato, T. *et al.* *Nature* **469**, 415–418 (2011).
6. Yilmaz, O. H. *et al.* *Nature* **486**, 490–495 (2012).
7. Kesse, E., Clavel-Chapelon, F. & Boutron-Ruault, M. *Am. J. Epidemiol.* **164**, 1085–1093 (2006).
8. Magee, J. A., Piskounova, E. & Morrison, S. J. *Cancer Cell* **21**, 283–296 (2012).
9. Nelson, E. R. *et al.* *Science* **342**, 1094–1098 (2013).
10. He, T.-C., Chan, T. A., Vogelstein, B. & Kinzler, K. W. *Cell* **99**, 335–345 (1999).
11. Gruber, L. *et al.* *PLoS ONE* **8**, e71661 (2013).

able to capture sufficient arrival events to draw statistically meaningful conclusions. At present, detector size and observation times are limited by the use of optical detectors<sup>2</sup>, which have areas of up to a few square kilometres and can operate only on clear, moonless nights.

The region between about  $10^{15.5}$  eV and  $10^{18.5}$  eV (the second energy value is called the ankle; Fig. 1) in the energy distribution of cosmic rays demarks an area of change. At energies higher than  $10^{15.5}$  eV, the composition of cosmic rays changes from lightweight particles, such as protons and helium nuclei, to increasingly heavy ones, such as the nuclei of carbon and heavier elements. These changes are thought to be associated with the maximum energies up to which Galactic sources — possibly supernova remnants (SNRs) — can accelerate cosmic rays<sup>3</sup>. The maximum energy for Galactic protons (the proton knee) is about  $10^{15.5}$  eV, whereas that for iron nuclei (the iron knee) is about  $10^{17}$  eV. The ankle is thought to be dominated by particles from extragalactic sources<sup>4,5</sup>, for example active galactic nuclei,  $\gamma$ -ray bursts or neutron stars. But, within this picture, the origin of cosmic rays between the iron knee and the ankle is unclear.

Buitink *et al.* investigated the cosmic-ray composition in this energy region using a new



decrease in Paneth-cell numbers. For instance, several signalling proteins (such as Jag1 and Jag2) that are normally produced by Paneth cells are upregulated in HFD-derived ISCs, sustaining niche-independent growth.

The incidence of human colorectal cancer correlates with diet-induced obesity<sup>7</sup>. Furthermore, adult stem cells are speculated to be the origin of some cancers<sup>8</sup>. Beyaz *et al.* showed that the increased pool of ISCs and ISC-like progenitors induced by an HFD predisposed mice to intestinal tumours. By contrast, calorie restriction — which also increases ISC numbers<sup>6</sup> — is associated with reduced tumour initiation<sup>1</sup>. The mechanistic differences underlying altered stem-cell function in each condition may partially explain this discrepancy. Calorie restriction is associated with increased interactions between the niche and ISCs, whereas HFD-associated, niche-independent growth allows stem cells to escape physiological regulation. How the molecular pathways modulated by these two dietary regimens intersect and communicate in ISCs remains to be investigated, and might identify putative therapeutic targets.

Gene-expression profiles often provide clues to the state of a cell. Nuclear receptors such as PPAR and LXR proteins sense nutrients and regulate gene expression, providing a connection between diet-induced metabolic changes and these profiles<sup>9</sup>. Beyaz and colleagues analysed gene expression and found that genetic targets of one PPAR, PPAR- $\delta$ , were upregulated in ISCs in HFD mice compared to ISCs in control mice. PPAR- $\delta$  is linked to a signalling cascade called the Wnt- $\beta$ -catenin pathway<sup>10</sup>, which is involved in the development of intestinal tumours. In the current study, genetic and pharmacological experiments revealed that activation of PPAR- $\delta$  and Wnt- $\beta$ -catenin signalling mediates, at least in part, the effects of dietary fats on ISC and progenitor function and intestinal tumour formation.

This work provides a plausible cellular and molecular explanation for how an excess of dietary fat remodels the intestine. However, questions remain about how the basic mechanisms of action of dietary fat affect systemic energy metabolism and other gastrointestinal diseases apart from cancer. For instance, it is known that diet can influence immune and metabolic activities by directly modulating the diversity and functions of the gut microbiota (the population of microorganisms that inhabit the gut), but it remains unclear whether and how such changes in the microbiota integrate with the PPAR- $\delta$  pathway. Because the microbiota differs between individuals, the interplay between microbiota and ISCs under HFD conditions might modulate an individual's tumour risk.

It would be interesting to investigate the contribution of ISCs to gut inflammatory disorders such as Crohn's disease because, as with cancer, an HFD accelerates the progress

of these disorders independently of obesity<sup>11</sup>. Furthermore, it is not known whether an HFD affects the gut's neuroendocrine system (the hormone-releasing cells that receive input from neurons), perhaps through its effects on ISCs and progenitors, to contribute to the metabolic alterations associated with obesity, type 2 diabetes or cardiovascular diseases.

The current study does not address whether the effects of an HFD on gut architecture are reversible. Moreover, it is unclear how changes in dietary regimens affect ISC function. Finally, further investigation will be needed to determine whether dietary or pharmacological interventions that target ISCs could maintain healthy intestinal function and reduce the incidence of tumours or other HFD-associated human diseases. Such research, building on the foundation provided by the current study, will be important for defining future steps in personalized human nutrition and health. ■

## SPACE SCIENCE

# Cosmic rays beyond the knees

**The development of a radio technique for detecting cosmic rays casts fresh light on the origins of some of these accelerated particles, and suggests that they might have travelled much farther than was previously thought. SEE LETTER P.70**

ANDREW M. TAYLOR

A technique for measuring the composition of cosmic rays in the energy range  $10^{17}$  to  $10^{17.5}$  electronvolts — a range thought to mark the transition point from Galactic to extragalactic cosmic rays — is reported by Buitink *et al.*<sup>1</sup> on page 70 of this issue. The findings have implications for our understanding of the sources of these mysterious rays.

Energetic protons and atomic nuclei arriving at Earth are classified as cosmic rays. The spread in energy of these particles is impressive, covering ten orders of magnitude (Fig. 1). The most energetic cosmic rays have energies more than 10 million times those achievable for protons at the Large Hadron Collider, the world's most powerful particle accelerator at CERN in Geneva, Switzerland. This begs the question of what cosmic sources can accelerate particles to such high energies.

Clues about the origin of cosmic rays come from both their composition and their energy spectra. Unfortunately, these particles arrive at Earth rather infrequently — only one particle with an energy greater than  $10^{17}$  eV arrives each day in every 10,000 square metres. This scarcity makes it difficult to build detectors

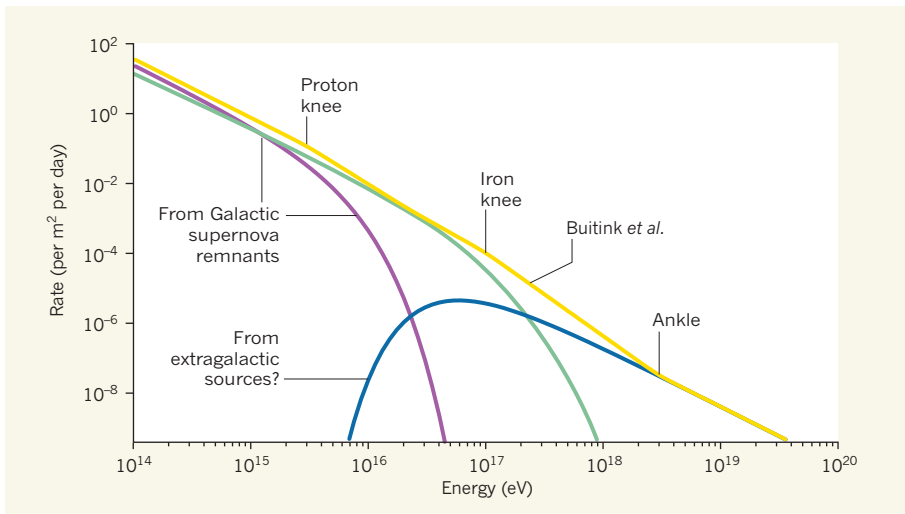
**Chi Luo and Pere Puigserver** are in the Department of Cell Biology, Harvard Medical School, and in the Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. e-mail: pere\_puigserver@dfci.harvard.edu

1. Nihaylova, M. M., Sabatini, D. M. & Yilmaz, O. H. *Cell Stem Cell* **14**, 292–305 (2014).
2. Nakada, D., Levi, B. P. & Morrison, S. J. *Neuron* **70**, 703–718 (2011).
3. Shyih-Chang, N., Daley, G. Q. & Cantley, L. C. *Development* **140**, 2535–2547 (2013).
4. Beyaz, S. *et al.* *Nature* **531**, 53–58 (2016).
5. Sato, T. *et al.* *Nature* **469**, 415–418 (2011).
6. Yilmaz, O. H. *et al.* *Nature* **486**, 490–495 (2012).
7. Kesse, E., Clavel-Chapelon, F. & Boutron-Ruault, M. *Am. J. Epidemiol.* **164**, 1085–1093 (2006).
8. Magee, J. A., Piskounova, E. & Morrison, S. J. *Cancer Cell* **21**, 283–296 (2012).
9. Nelson, E. R. *et al.* *Science* **342**, 1094–1098 (2013).
10. He, T.-C., Chan, T. A., Vogelstein, B. & Kinzler, K. W. *Cell* **99**, 335–345 (1999).
11. Gruber, L. *et al.* *PLoS ONE* **8**, e71661 (2013).

able to capture sufficient arrival events to draw statistically meaningful conclusions. At present, detector size and observation times are limited by the use of optical detectors<sup>2</sup>, which have areas of up to a few square kilometres and can operate only on clear, moonless nights.

The region between about  $10^{15.5}$  eV and  $10^{18.5}$  eV (the second energy value is called the ankle; Fig. 1) in the energy distribution of cosmic rays demarks an area of change. At energies higher than  $10^{15.5}$  eV, the composition of cosmic rays changes from lightweight particles, such as protons and helium nuclei, to increasingly heavy ones, such as the nuclei of carbon and heavier elements. These changes are thought to be associated with the maximum energies up to which Galactic sources — possibly supernova remnants (SNRs) — can accelerate cosmic rays<sup>3</sup>. The maximum energy for Galactic protons (the proton knee) is about  $10^{15.5}$  eV, whereas that for iron nuclei (the iron knee) is about  $10^{17}$  eV. The ankle is thought to be dominated by particles from extragalactic sources<sup>4,5</sup>, for example active galactic nuclei,  $\gamma$ -ray bursts or neutron stars. But, within this picture, the origin of cosmic rays between the iron knee and the ankle is unclear.

Buitink *et al.* investigated the cosmic-ray composition in this energy region using a new



**Figure 1 | The energy spectrum of cosmic rays.** When the number of cosmic rays per square metre that arrive at Earth's surface per day is plotted on a logarithmic scale against the energies of the cosmic rays, also on a logarithmic scale, the overall graph (yellow line) between  $10^{14}$  and  $10^{19}$  electronvolts contains four distinct linear regions, connected by changes in slope called the proton knee, the iron knee and the ankle. The slope changes correspond to transitions between the composition or sources of cosmic rays that dominate at different energies. Buitink *et al.*<sup>1</sup> report on observations of cosmic rays between the iron knee and the ankle, providing clues about their origins.

radio-based instrument. This can detect more cosmic rays in a given period than can optical detectors of equivalent effective area, because it can operate both day and night. The authors' detection method relies on descriptions<sup>6,7</sup> of the radio signals generated by air showers — the cascades of energetic charged particles and electromagnetic radiation produced when cosmic rays enter the atmosphere. The radio signal is produced by the interaction of the charged particles with Earth's magnetic field, and by the development of a charge imbalance within the shower through a phenomenon called the Askarayan effect<sup>8</sup>; the contributions of these two effects to the signal are of comparable magnitude.

This knowledge allowed Buitink and colleagues to probe the air-shower profile produced by cosmic rays in the energy range of interest. Air showers involve a chain of processes that rapidly increase the number of energetic particles (known as secondary particles) within them. The rate of transfer of a cosmic ray's energy to secondary particles depends on its nuclear composition. The authors could discriminate between different species of cosmic ray — that is, the type of particle that was originally accelerated — from the depth in the atmosphere at which the maximum number of air-shower particles occurred.

From a sample of 118 such profile measurements, the researchers concluded that the fraction of cosmic rays consisting of protons and helium nuclei in the energy band  $10^{17}$  to  $10^{17.5}$  eV is between 38% and 98%, with a 99% confidence level; the best-fit value for this light-mass fraction is about 80%. These results are consistent with previous results attained using optical techniques<sup>2,9</sup>, but this is the first

time that such a composition measurement has been made using a radio instrument.

The success of Buitink and co-workers' method provides fresh clues about the origin of cosmic rays between the iron knee and the ankle. It could be that cosmic rays in this energy range are associated with an extragalactic component, which would challenge the idea

that the ankle represents the onset of this component in the energy distribution. Alternatively, these cosmic rays might have a Galactic origin. This would indicate the existence of a second population of Galactic sources capable of accelerating particles to considerably higher energies than those achievable by SNRs. Either way, the authors' findings support the idea of light-mass cosmic rays in the knee-to-ankle region that must now be explained.

The powerful method used to make this measurement might also seed a new era of cosmic-ray science, in which previously unknown spectral features and changes in composition at energies within the transition zone are probed with unprecedented detail. Perhaps in this way we will uncover further surprises and evidence of extragalactic sources. ■

**Andrew M. Taylor** is in the High Energy Astrophysics Group, School of Cosmic Physics, Dublin Institute for Advanced Studies, Dublin 2, Ireland.  
e-mail: [taylora@cp.dias.ie](mailto:taylora@cp.dias.ie)

1. Buitink, S. *et al.* *Nature* **531**, 70–73 (2016).
2. Berezhnev, S. F. *et al.* *Proc. 32nd ICRC Beijing* **1**, 209–212 (11–18 August 2011).
3. Bell, A. R., Schure, K. M., Reville, B. & Giacinti, G. *Mon. Not. R. Astron. Soc.* **431**, 415–429 (2013).
4. Hillas, A. M. *Annu. Rev. Astron. Astrophys.* **22**, 425–444 (1984).
5. Waxman, E. *Phys. Rev. Lett.* **75**, 386–389 (1995).
6. Huege, T., Ludwig, M. & James, C. W. *AIP Conf. Proc.* **1535**, 128 (2013).
7. Alvarez-Muñiz, J., Carvalho, W. R. Jr & Zas, E. *Astropart. Phys.* **35**, 325–341 (2012).
8. Askarayan, G. *Sov. Phys. JETP* **14**, 441–443 (1962).
9. Abu-Zayyad, T. *et al.* *Astrophys. J.* **557**, 686 (2001).

## REGENERATION

# Stem cells make the bowel nervous

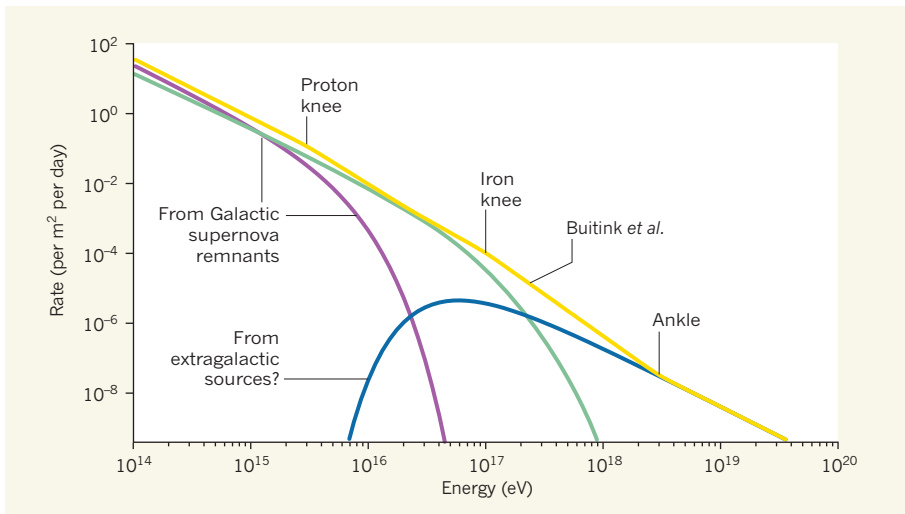
**In Hirschsprung disease, the enteric nervous system (ENS) is missing from the distal bowel. It emerges that postnatal transplantation of stem-cell-derived ENS precursors can prevent death in a mouse model of the disease. SEE LETTER P.105**

**ROBERT O. HEUCKEROOTH**

**T**he enteric nervous system (ENS) is a network of neurons and supporting glial cells in the bowel wall that is essential for digestion<sup>1</sup>. When ENS precursor cells fail to migrate through the full length of the bowel during the first trimester of pregnancy, a life-threatening birth defect called Hirschsprung disease ensues — a prominent symptom of which is constant contraction of the affected bowel regions<sup>2,3</sup>. The standard treatment for children with Hirschsprung disease is removal of the abnormal bowel, but many children continue to have bowel problems after surgery<sup>4</sup>. On

page 105 of this issue, Fattahi *et al.*<sup>5</sup> describe a method for generating ENS precursors from stem cells. Remarkably, transplantation of these cells into an animal model of Hirschsprung disease prevented premature death.

The ENS contains about as many neurons as the spinal cord, and its diversity of neuronal subtypes rivals that of the brain<sup>6</sup>. This complexity allows the ENS to recognize sensory input from both the bowel wall and within the bowel, and to produce integrated bowel motility patterns that facilitate nutrient absorption<sup>7</sup>. The ENS also influences bowel inflammatory cells, blood vessels, smooth muscle, intestinal pacemakers and the



**Figure 1 | The energy spectrum of cosmic rays.** When the number of cosmic rays per square metre that arrive at Earth's surface per day is plotted on a logarithmic scale against the energies of the cosmic rays, also on a logarithmic scale, the overall graph (yellow line) between  $10^{14}$  and  $10^{19}$  electronvolts contains four distinct linear regions, connected by changes in slope called the proton knee, the iron knee and the ankle. The slope changes correspond to transitions between the composition or sources of cosmic rays that dominate at different energies. Buitink *et al.*<sup>1</sup> report on observations of cosmic rays between the iron knee and the ankle, providing clues about their origins.

radio-based instrument. This can detect more cosmic rays in a given period than can optical detectors of equivalent effective area, because it can operate both day and night. The authors' detection method relies on descriptions<sup>6,7</sup> of the radio signals generated by air showers — the cascades of energetic charged particles and electromagnetic radiation produced when cosmic rays enter the atmosphere. The radio signal is produced by the interaction of the charged particles with Earth's magnetic field, and by the development of a charge imbalance within the shower through a phenomenon called the Askarayan effect<sup>8</sup>; the contributions of these two effects to the signal are of comparable magnitude.

This knowledge allowed Buitink and colleagues to probe the air-shower profile produced by cosmic rays in the energy range of interest. Air showers involve a chain of processes that rapidly increase the number of energetic particles (known as secondary particles) within them. The rate of transfer of a cosmic ray's energy to secondary particles depends on its nuclear composition. The authors could discriminate between different species of cosmic ray — that is, the type of particle that was originally accelerated — from the depth in the atmosphere at which the maximum number of air-shower particles occurred.

From a sample of 118 such profile measurements, the researchers concluded that the fraction of cosmic rays consisting of protons and helium nuclei in the energy band  $10^{17}$  to  $10^{17.5}$  eV is between 38% and 98%, with a 99% confidence level; the best-fit value for this light-mass fraction is about 80%. These results are consistent with previous results attained using optical techniques<sup>2,9</sup>, but this is the first

time that such a composition measurement has been made using a radio instrument.

The success of Buitink and co-workers' method provides fresh clues about the origin of cosmic rays between the iron knee and the ankle. It could be that cosmic rays in this energy range are associated with an extragalactic component, which would challenge the idea

that the ankle represents the onset of this component in the energy distribution. Alternatively, these cosmic rays might have a Galactic origin. This would indicate the existence of a second population of Galactic sources capable of accelerating particles to considerably higher energies than those achievable by SNRs. Either way, the authors' findings support the idea of light-mass cosmic rays in the knee-to-ankle region that must now be explained.

The powerful method used to make this measurement might also seed a new era of cosmic-ray science, in which previously unknown spectral features and changes in composition at energies within the transition zone are probed with unprecedented detail. Perhaps in this way we will uncover further surprises and evidence of extragalactic sources. ■

**Andrew M. Taylor** is in the High Energy Astrophysics Group, School of Cosmic Physics, Dublin Institute for Advanced Studies, Dublin 2, Ireland.  
e-mail: [taylora@cp.dias.ie](mailto:taylora@cp.dias.ie)

1. Buitink, S. *et al.* *Nature* **531**, 70–73 (2016).
2. Berezhnev, S. F. *et al.* *Proc. 32nd ICRC Beijing* **1**, 209–212 (11–18 August 2011).
3. Bell, A. R., Schure, K. M., Reville, B. & Giacinti, G. *Mon. Not. R. Astron. Soc.* **431**, 415–429 (2013).
4. Hillas, A. M. *Annu. Rev. Astron. Astrophys.* **22**, 425–444 (1984).
5. Waxman, E. *Phys. Rev. Lett.* **75**, 386–389 (1995).
6. Huege, T., Ludwig, M. & James, C. W. *AIP Conf. Proc.* **1535**, 128 (2013).
7. Alvarez-Muñiz, J., Carvalho, W. R. Jr & Zas, E. *Astropart. Phys.* **35**, 325–341 (2012).
8. Askarayan, G. *Sov. Phys. JETP* **14**, 441–443 (1962).
9. Abu-Zayyad, T. *et al.* *Astrophys. J.* **557**, 686 (2001).

## REGENERATION

# Stem cells make the bowel nervous

**In Hirschsprung disease, the enteric nervous system (ENS) is missing from the distal bowel. It emerges that postnatal transplantation of stem-cell-derived ENS precursors can prevent death in a mouse model of the disease. SEE LETTER P.105**

**ROBERT O. HEUCKEROOTH**

**T**he enteric nervous system (ENS) is a network of neurons and supporting glial cells in the bowel wall that is essential for digestion<sup>1</sup>. When ENS precursor cells fail to migrate through the full length of the bowel during the first trimester of pregnancy, a life-threatening birth defect called Hirschsprung disease ensues — a prominent symptom of which is constant contraction of the affected bowel regions<sup>2,3</sup>. The standard treatment for children with Hirschsprung disease is removal of the abnormal bowel, but many children continue to have bowel problems after surgery<sup>4</sup>. On

page 105 of this issue, Fattahi *et al.*<sup>5</sup> describe a method for generating ENS precursors from stem cells. Remarkably, transplantation of these cells into an animal model of Hirschsprung disease prevented premature death.

The ENS contains about as many neurons as the spinal cord, and its diversity of neuronal subtypes rivals that of the brain<sup>6</sup>. This complexity allows the ENS to recognize sensory input from both the bowel wall and within the bowel, and to produce integrated bowel motility patterns that facilitate nutrient absorption<sup>7</sup>. The ENS also influences bowel inflammatory cells, blood vessels, smooth muscle, intestinal pacemakers and the



epithelial cells that line the bowel.

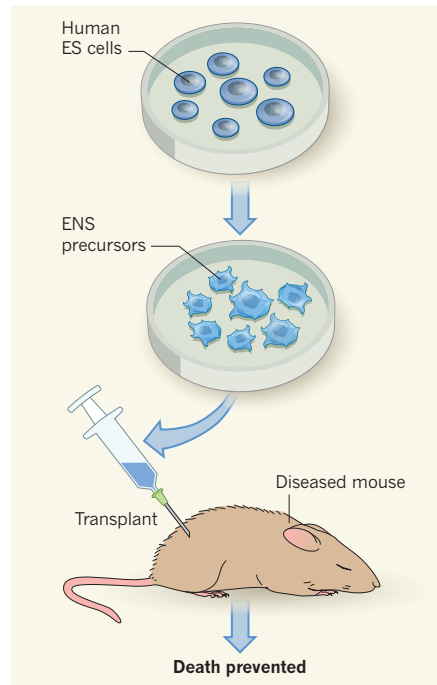
It is therefore not surprising that children with Hirschsprung disease develop abdominal distension, vomiting and constipation, fail to grow normally and can die from sepsis (a bacterial infection of the bloodstream). At least one-third<sup>3</sup> of children with Hirschsprung disease continue to have serious problems after surgery, including a life-threatening syndrome called enterocolitis. Furthermore, some children with this disease have so little bowel that is innervated by the ENS that they require intravenous nutrient delivery to survive.

Exciting work<sup>8</sup> suggests that regenerative medicine could one day offer an alternative to surgery for treating Hirschsprung disease. In this approach, stem cells would be transplanted into and would restore function in bowel regions in which the ENS is missing. Ideally, the transplanted cells would come from the affected child (known as autologous transplantation) to avoid immune rejection, and non-surgical methods would be used for cell delivery.

Of particular interest for this type of therapy are gut-derived ENS stem cells, which can be isolated from the human bowel at all ages and cultured *in vitro*. Following culture, these stem cells can be reimplanted in the bowel wall. They then migrate to the normal site of the ENS and differentiate into neurons and glial cells that mimic those of the native ENS<sup>9</sup>. However, this therapeutic approach faces several challenges, including difficulty producing enough gut-derived cells, limited cell migration, limited data about long-term safety and minimal information about the ability of these cells to restore gut function.

Fattahi *et al.* address some of these problems using human embryonic stem (ES) cells, which are derived from early embryos and can give rise to every cell type in the body. To direct differentiation of human ES cells towards an ENS precursor lineage, the authors modulated signalling pathways that control development by inhibiting SMAD and glycogen synthase kinase proteins, and then treated the cells with the metabolite retinoic acid. Under these conditions, human ES cells differentiated into cells that resemble ENS precursors from the vagal region of the developing spinal cord (called the vagal neural crest)<sup>10</sup>. The authors refer to these cells as enteric neural-crest (ENC) precursors.

These ENC precursors shared several key features with ENS precursors. For instance, when transplanted into the vagal neural-crest regions of developing chick embryos, ENC precursors often migrated to the bowel, like normal ENS precursors. When transplanted to the colon of young mice, ENC precursors populated the bowel close to the location of the normal ENS, but migrated even more quickly than fetal ENS-derived cells. When grown alongside human ES-cell-derived smooth-muscle cells, ENC precursors enhanced muscle differentiation and became neurons that could induce



**Figure 1 | Help for a model of Hirschsprung disease.** Hirschsprung disease is a birth defect in which the enteric nervous system (ENS) fails to colonize the full length of the bowel during early pregnancy. To investigate possible therapies for the disease, Fattahi *et al.*<sup>5</sup> grew human embryonic stem (ES) cells (which can give rise to every cell type of the body) *in vitro* under conditions that encouraged them to differentiate into cells resembling ENS precursors. The authors transplanted these cultured cells into the colons of mice with a genetic mutation that causes a Hirschsprung-like disease. The transplant prevented premature death in these mice, although how the cells achieved this feat is not clear.

muscle contraction when activated. Following an extended period of *in vitro* culture with vitamin C and the growth factor GDNF, ENC precursors produced diverse neuronal and glial cells similar to those of the ENS. Most impressively, when ENC precursors were transplanted into the colons of mice with Hirschsprung-like disease, survival rates improved dramatically over a short time interval (Fig. 1).

Finally, Fattahi *et al.* used ENC precursors harbouring a genetic mutation that predisposes humans to Hirschsprung disease to perform an *in vitro* drug screen, and discovered that inhibiting the protease enzyme BACE2 enhanced ENC-precursor migration. The gene that encodes BACE2 is located in a region of chromosome 21 whose duplication increases the risk of Hirschsprung disease. This finding may be relevant to Down syndrome, in which children are born with three copies of chromosome 21 and rates of Hirschsprung disease are increased by as much as 100-fold<sup>3</sup>.

This study establishes a potentially limitless source of cells similar to those of the vagal neural crest that could be tested for use in transplants to treat children with Hirschsprung

disease or other disorders in which the ENS is defective. In an ideal therapy, ENC precursors would be produced from ‘induced pluripotent’ stem cells, which closely resemble human ES cells<sup>11</sup>, but can be derived from the skin or blood cells of affected children, removing the need for embryo-derived cells and post-transplant immunosuppression. Fattahi and colleagues provide preliminary data to suggest that this strategy will work well. Furthermore, the human ES-cell-derived ENC precursors they produced migrate efficiently through the bowel and could potentially be delivered through an endoscope, avoiding invasive surgery.

Although these advances are exciting, many questions remain. In particular, it is unlikely that transplanted ENC precursors recreated a normal ENS in the Hirschsprung model mice, given the rapidity with which transplantation rescued lethal bowel disease. Instead, minimally organized ENC precursors might have modulated immune activity or enhanced epithelial-cell function and repair by releasing neurotransmitter molecules (or other factors). Identifying these ENC-precursor-derived factors might lead to the development of other treatment or prevention strategies that obviate the need for cell-based therapies. Similarly, BACE2 targets that influence ENC-precursor migration could be used to enhance stem-cell therapy or to prevent Hirschsprung disease.

Finally, the effect of transplanted ENC precursors on bowel motility and long-term safety needs to be addressed. Nonetheless, Fattahi and colleagues’ study moves us one step closer to a time when autologous stem-cell therapy could replace surgery as a primary treatment for children with Hirschsprung disease. ■

**Robert O. Heuckeroth** is at the Children’s Hospital of Philadelphia Research Institute and the Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.  
e-mail: heuckeroth@email.chop.edu

- Kunze, W. A. A. & Furness, J. B. *Annu. Rev. Physiol.* **61**, 117–142 (1999).
- McKeown, S. J., Stamp, L., Hao, M. M. & Young, H. M. *Wiley Interdisc. Rev. Dev. Biol.* **2**, 113–129 (2013).
- Heuckeroth, R. O. in *Pediatric Neurogastroenterology: Gastrointestinal Motility and Functional Disorders in Children* (eds Faure, C., Di Lorenzo, C. & Thapar, N.) 271–283 (Springer, 2013).
- Burkhardt, D. D., Graham, J. M., Short, S. S. & Frykman, P. K. *Clin. Pediatr.* **53**, 71–81 (2014).
- Fattahi, F. *et al. Nature* **531**, 105–109 (2016).
- Furness, J. B. *Nature Rev. Gastroenterol. Hepatol.* **9**, 286–294 (2012).
- Neunlist, M. & Schemmann, M. *J. Physiol. (Lond.)* **592**, 2959–2965 (2014).
- Burns, A. J. & Thapar, N. *Nature Rev. Gastroenterol. Hepatol.* **11**, 317–328 (2014).
- Metzger, M., Caldwell, C., Barlow, A. J., Burns, A. J. & Thapar, N. *Gastroenterology* **136**, 2214–2225 (2009).
- Avetisyan, M., Schill, E. M. & Heuckeroth, R. O. *J. Clin. Invest.* **125**, 899–907 (2015).
- Takahashi, K. & Yamanaka, S. *Development* **142**, 3274–3285 (2015).

This article was published online on 10 February 2016.

# Genomic analyses identify molecular subtypes of pancreatic cancer

Peter Bailey<sup>1,2</sup>, David K. Chang<sup>2,3,4,5</sup>, Katia Nones<sup>1,6</sup>, Amber L. Johns<sup>3</sup>, Ann-Marie Patch<sup>1,6</sup>, Marie-Claude Gingras<sup>7,8,9</sup>, David K. Miller<sup>1,3</sup>, Angelika N. Christ<sup>1</sup>, Tim J. C. Bruxner<sup>1</sup>, Michael C. Quinn<sup>1,6</sup>, Craig Nourse<sup>1,2</sup>, L. Charles Murtaugh<sup>10</sup>, Ivon Harliwong<sup>1</sup>, Senel Idrisoglu<sup>1</sup>, Suzanne Manning<sup>1</sup>, Ehsan Nourbakhsh<sup>1</sup>, Shivangi Wani<sup>1,6</sup>, Lynn Fink<sup>1</sup>, Oliver Holmes<sup>1,6</sup>, Venessa Chin<sup>3</sup>, Matthew J. Anderson<sup>1</sup>, Stephen Kazakoff<sup>1,6</sup>, Conrad Leonard<sup>1,6</sup>, Felicity Newell<sup>1</sup>, Nick Waddell<sup>1</sup>, Scott Wood<sup>1,6</sup>, Qinying Xu<sup>1,6</sup>, Peter J. Wilson<sup>1</sup>, Nicole Cloonan<sup>1,6</sup>, Karin S. Kassahn<sup>1,11,12</sup>, Darrin Taylor<sup>1</sup>, Kelly Quek<sup>1</sup>, Alan Robertson<sup>1</sup>, Lorena Pantano<sup>13</sup>, Laura Mincarelli<sup>2</sup>, Luis N. Sanchez<sup>2</sup>, Lisa Evers<sup>2</sup>, Jianmin Wu<sup>3</sup>, Mark Pinese<sup>3</sup>, Mark J. Cowley<sup>3</sup>, Marc D. Jones<sup>2,3</sup>, Emily K. Colvin<sup>3</sup>, Adnan M. Nagrial<sup>3</sup>, Emily S. Humphrey<sup>3</sup>, Lorraine A. Chantrill<sup>3,14</sup>, Amanda Mawson<sup>3</sup>, Jeremy Humphris<sup>3</sup>, Angela Chou<sup>3,15</sup>, Marina Pajic<sup>3,16</sup>, Christopher J. Scarlett<sup>3,17</sup>, Andreia V. Pinho<sup>3</sup>, Marc Giry-Laterriere<sup>3</sup>, Ilse Rooman<sup>3</sup>, Jaswinder S. Samra<sup>18,19</sup>, James G. Kench<sup>3,19,20</sup>, Jessica A. Lovell<sup>3</sup>, Neil D. Merrett<sup>5,21</sup>, Christopher W. Toon<sup>3</sup>, Krishna Epari<sup>22</sup>, Nam Q. Nguyen<sup>23</sup>, Andrew Barbour<sup>24</sup>, Nikolajs Zeps<sup>25</sup>, Kim Moran-Jones<sup>2</sup>, Nigel B. Jamieson<sup>2,26,27</sup>, Janet S. Graham<sup>2,28</sup>, Fraser Duthie<sup>29</sup>, Karin Oien<sup>3,29</sup>, Jane Hair<sup>30</sup>, Robert Grützmann<sup>31†</sup>, Anirban Maitra<sup>32</sup>, Christine A. Iacobuzio-Donahue<sup>33</sup>, Christopher L. Wolfgang<sup>34,35</sup>, Richard A. Morgan<sup>34</sup>, Rita T. Lawlor<sup>36,37</sup>, Vincenzo Corbo<sup>36</sup>, Claudio Bassi<sup>38</sup>, Borislav Rusev<sup>36</sup>, Paola Capelli<sup>37</sup>, Roberto Salvia<sup>38</sup>, Giampaolo Tortora<sup>39</sup>, Debabrata Mukhopadhyay<sup>40</sup>, Gloria M. Petersen<sup>40</sup>, Australian Pancreatic Cancer Genome Initiative\*, Donna M. Munzy<sup>7,8</sup>, William E. Fisher<sup>41</sup>, Saadia A. Karim<sup>42</sup>, James R. Eshleman<sup>34</sup>, Ralph H. Hruban<sup>34</sup>, Christian Pilarsky<sup>31</sup>, Jennifer P. Morton<sup>42</sup>, Owen J. Sansom<sup>42,43</sup>, Aldo Scarpa<sup>36,37</sup>, Elizabeth A. Musgrove<sup>2</sup>, Ulla-Maja Hagbo Bailey<sup>2</sup>, Oliver Hofmann<sup>2,13</sup>, Robert L. Sutherland<sup>3‡</sup>, David A. Wheeler<sup>7,8</sup>, Anthony J. Gill<sup>3,19</sup>, Richard A. Gibbs<sup>7,8</sup>, John V. Pearson<sup>1,6</sup>, Nicola Waddell<sup>1,6</sup>, Andrew V. Biankin<sup>2,3,4,5,27</sup> & Sean M. Grimmond<sup>1,2,44</sup>

**Integrated genomic analysis of 456 pancreatic ductal adenocarcinomas identified 32 recurrently mutated genes that aggregate into 10 pathways: KRAS, TGF- $\beta$ , WNT, NOTCH, ROBO/SLIT signalling, G1/S transition, SWI-SNF, chromatin modification, DNA repair and RNA processing. Expression analysis defined 4 subtypes: (1) squamous; (2) pancreatic progenitor; (3) immunogenic; and (4) aberrantly differentiated endocrine exocrine (ADEX) that correlate with histopathological characteristics. Squamous tumours are enriched for TP53 and KDM6A mutations, upregulation of the TP63 $\Delta$ N transcriptional network, hypermethylation of pancreatic endodermal cell-fate determining genes and have a poor prognosis. Pancreatic progenitor tumours preferentially express genes involved in early pancreatic development (FOXA2/3, PDX1 and MNX1). ADEX tumours displayed upregulation of genes that regulate networks involved in KRAS activation, exocrine (NR5A2 and RBPL), and endocrine differentiation (NEUROD1 and NKX2-2). Immunogenic tumours contained upregulated immune networks including pathways involved in acquired immune suppression. These data infer differences in the molecular evolution of pancreatic cancer subtypes and identify opportunities for therapeutic development.**

<sup>1</sup>Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia. <sup>2</sup>Wolfson Wohl Cancer Research Centre, Institute of Cancer Sciences, University of Glasgow, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK. <sup>3</sup>The Kinghorn Cancer Centre, 370 Victoria St, Darlinghurst, and the Cancer Research Program, Garvan Institute of Medical Research, 384 Victoria St, Darlinghurst, Sydney, New South Wales 2010, Australia. <sup>4</sup>Department of Surgery, Bankstown Hospital, Eldridge Road, Bankstown, Sydney, New South Wales 2200, Australia. <sup>5</sup>South Western Sydney Clinical School, Faculty of Medicine, University of New South Wales, Liverpool, New South Wales 2170, Australia. <sup>6</sup>QIMR Berghofer Medical Research Institute, Herston, Queensland 4006, Australia. <sup>7</sup>Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>8</sup>Michael DeBakey Department of Surgery, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>9</sup>Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>10</sup>Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA. <sup>11</sup>Genetic and Molecular Pathology, SA Pathology, Adelaide, South Australia 5000, Australia. <sup>12</sup>School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5000, Australia. <sup>13</sup>Harvard Chan Bioinformatics Core, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, USA. <sup>14</sup>Macarthur Cancer Therapy Centre, Campbelltown Hospital, New South Wales 2560, Australia. <sup>15</sup>Department of Pathology, SydPath, St Vincent's Hospital, Sydney, NSW 2010, Australia. <sup>16</sup>St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, New South Wales 2052, Australia. <sup>17</sup>School of Environmental & Life Sciences, University of Newcastle, Ourimbah, New South Wales 2258, Australia. <sup>18</sup>Department of Surgery, Royal North Shore Hospital, St Leonards, Sydney, New South Wales 2065, Australia. <sup>19</sup>University of Sydney, Sydney, New South Wales 2006, Australia. <sup>20</sup>Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown New South Wales 2050, Australia. <sup>21</sup>School of Medicine, University of Western Sydney, Penrith, New South Wales 2175, Australia. <sup>22</sup>Fiona Stanley Hospital, Robin Warren Drive, Murdoch, Western Australia 6150, Australia. <sup>23</sup>Department of Gastroenterology, Royal Adelaide Hospital, North Terrace, Adelaide, South Australia 5000, Australia. <sup>24</sup>Department of Surgery, Princess Alexandra Hospital, Ipswich Rd, Woollongabba, Queensland 4102, Australia. <sup>25</sup>School of Surgery M507, University of Western Australia, 35 Stirling Hwy, Nedlands 6009, Australia and St John of God Pathology, 12 Salvado Rd, Subiaco, Western Australia 6008, Australia. <sup>26</sup>Academic Unit of Surgery, School of Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow Royal Infirmary, Glasgow G4 0SF, UK. <sup>27</sup>West of Scotland Pancreatic Unit, Glasgow Royal Infirmary, Glasgow G31 2ER, UK. <sup>28</sup>Department of Medical Oncology, Beatson West of Scotland Cancer Centre, 1053 Great Western Road, Glasgow G12 0YN, UK. <sup>29</sup>Department of Pathology, Southern General Hospital, Greater Glasgow & Clyde NHS, Glasgow G51 4TF, UK. <sup>30</sup>GGC Bio-repository, Pathology Department, Southern General Hospital, 1345 Govan Road, Glasgow G51 4TY, UK. <sup>31</sup>Department of Surgery, TU Dresden, Fetscherstr. 74, 01307 Dresden, Germany. <sup>32</sup>Departments of Pathology and Translational Molecular Pathology, UT MD Anderson Cancer Center, Houston Texas 77030, USA. <sup>33</sup>The David M. Rubenstein Pancreatic Cancer Research Center and Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. <sup>34</sup>Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA. <sup>35</sup>Department of Surgery, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA. <sup>36</sup>ARC-Net Applied Research on Cancer Centre, University and Hospital Trust of Verona, Verona 37134, Italy. <sup>37</sup>Department of Pathology and Diagnostics, University of Verona, Verona 37134, Italy. <sup>38</sup>Department of Surgery, Pancreas Institute, University and Hospital Trust of Verona, Verona 37134, Italy. <sup>39</sup>Department of Medical Oncology, Comprehensive Cancer Centre, University and Hospital Trust of Verona, Verona 37134, Italy. <sup>40</sup>Mayo Clinic, Rochester, Minnesota 55905, USA. <sup>41</sup>Elkins Pancreas Center, Baylor College of Medicine, One Baylor Plaza, MS226, Houston, Texas 77030-3411, USA. <sup>42</sup>Cancer Research UK Beatson Institute, Glasgow G61 1BD, UK. <sup>43</sup>Institute for Cancer Science, University of Glasgow, Glasgow G12 8QQ, UK. <sup>44</sup>University of Melbourne, Parkville, Victoria 3010, Australia.

†Present address: Universitätsklinikum Erlangen, Department of Surgery, 91054 Erlangen, Germany.

\*Lists of participants and their affiliations appear in the Supplementary Information.

‡Deceased.

Pancreatic cancer (PC) is the fourth leading cause of cancer death in Western societies, and projected to be the second within a decade<sup>1</sup>. It has a median survival measured in months and a five-year survival of <5%. Advances in therapy have only achieved incremental improvements in overall outcome, but can provide notable benefit for undefined subgroups of patients. As a consequence, there is an urgent need to better understand the molecular pathology of PC in order to improve patient selection for current treatment options, and to develop novel therapeutic strategies.

Genomic analyses of pancreatic cancer reveal a complex mutational landscape with four common oncogenic events in well-known cancer genes (*KRAS*, *TP53*, *SMAD4* and *CDKN2A*), amongst a milieu of genes mutated at low prevalence. Despite this heterogeneity, oncogenic point mutations of individual genes aggregate into core molecular pathways including DNA damage repair, cell cycle regulation, TGF- $\beta$  signalling, chromatin regulation and axonal guidance<sup>2–5</sup>. Increasingly sophisticated analyses are revealing biologically important events with clinical significance, including whole-genome sequencing, which sub-classifies PC into 4 subtypes based on the frequency and distribution of structural variation. Those termed unstable due to a large number of structural variants correlate with defects in DNA maintenance and therapeutic responsiveness to platinum based therapies<sup>2</sup>. Aberrations in other features that characterize cancer genomes, including mutational signatures<sup>6</sup>, and differential methylation<sup>7</sup> are providing deeper insights into disease pathophysiology.

Here we performed a comprehensive integrated genomic analysis of 456 PCs and their histopathological variants using a combination of whole-genome and deep-exome sequencing, with gene copy number analysis to determine the mutational mechanisms and candidate genomic events important in pancreatic carcinogenesis. RNA expression profiles were used to define four subtypes and the different transcriptional networks that underpin them. These subtypes are associated with distinct histopathological characteristics and differential survival. Genomic and epigenetic features that characterize each subtype infer different mechanisms of molecular evolution.

## Mutational landscape of PC

Study participants were recruited and consent for genomic sequencing obtained through the Australian Pancreatic Cancer Genome Initiative (APGI; <http://www.pancreaticcancer.net.au>) as part of the International Cancer Genome Consortium (ICGC; <http://www.icgc.org>). The 382 APGI group consisted of participants with primarily treatment-naïve resected PC, which were pancreatic ductal adenocarcinoma (PDAC) and its variants (adenosquamous, colloid, PDAC associated with intra-ductal papillary mucinous neoplasm (IPMN)) and a small number of rare acinar cell carcinomas (Supplementary Table 1). We detected 23,538 high confidence coding mutations<sup>2,8,9</sup>, of which, 7,377 were verified using orthogonal approaches (Supplementary Tables 1, 2 and 19). A total of 21,208 high confidence genomic rearrangements were also identified (Supplementary Tables 3 and 4)<sup>2,8</sup>. To maximize the power to define coding driver mutations, 74 previously published PC exomes<sup>3–5</sup> were included to yield a final cohort of 456 tumours. OncodriverFM detected 32 significantly mutated genes (false discovery rate (FDR)  $\leq 0.1$ ), 22 of which were also identified by MutsigCV2 ( $Q < 0.1$ ) and/or were supported by HOTNET2 analysis (Methods and Supplementary Table 5). These significantly mutated genes aggregated into 10 molecular mechanisms (Extended Data Fig. 1): with activating mutations of *KRAS* in 92%; disruption of G1/S checkpoint machinery (*TP53*, *CDKN2A* and *TP53BP2*) in 78%; TGF- $\beta$  signalling (*SMAD4*, *SMAD3*, *TGFBR1*, *TGFBR2*, *ACVR1B* and *ACVR2A*) in 47%; histone modification (*KDM6A*, *SETD2* and ASCOM complex members *MLL2* and *MLL3*) in 24%; the SWI/SNF complex (*ARID1A*, *PBRM1* and *SMARCA4*) in 14%; the BRCA pathway (*BRCA1*, *BRCA2*, *ATM* and *PALB2*: 5% germline, 12% somatic); WNT signalling defects through *RNF43* mutation (5%); and RNA processing genes, *SF3B1*, *U2AF1* and *RBM10* (16%). *RBM10* is implicated in lung cancer<sup>10</sup>, where inactivating mutations influence expression of oncogenic isoforms of *NUMB*<sup>11</sup>.

*SF3B1* mutations in PC were aggregated at the K700E mutation hotspot common in myelodysplastic syndrome, breast and lung cancer<sup>12</sup> and presents a potential therapeutic target<sup>13</sup>. Mutations in other genes encoding splicing machinery: *SF3A1*, *U2AF2*, *SF1* and *RBM6* were also identified (Extended Data Fig. 2 and Supplementary Table 6).

GISTIC2 identified 50 regions of recurrent gain (43 focal, 7 chromosomal arms) and 73 regions of loss (61 focal, 12 chromosomal arms) (Supplementary Tables 7–9). These regions included known oncogenes *MET*, *NOTCH1* and *GATA6* and tumour suppressor genes *CDKN2A*, *SMAD4*, *TP53*, *BRCA1*, *ARID1A*, *PBRM1* and *SMARCA4*. Integrating copy number and expression data identified a number of genes/amplicons implicated in the progression of other cancer types that exhibited concordant gene expression changes (Supplementary Table 10). These included: amplification of *MIB1*, a known mediator of NOTCH signalling and pancreas development<sup>14</sup> and the *CCNE1-UR11* amplicon at 19q12 (Extended Data Fig. 2b). *CCNE1* is a marker of poor prognosis in ovarian, breast and lung cancers and is associated with resistance to platinum based therapy<sup>15</sup>. Recent small interfering RNA (siRNA) screening of PC cell lines provides supportive evidence for *CCNE1* amplification as an important mechanism in pancreatic carcinogenesis, and may represent a therapeutic opportunity using CDK inhibitors<sup>16</sup>.

DNA deamination, ectopic APOBEC activity, BRCA-deficiency and mismatch repair were re-affirmed as the predominant mutational mechanisms in PC. Chromothrptic and break-fusion-bridge related genomic catastrophes were uncommon (12%; Supplementary Table 11). Somatic LINE-1 retro-transposition of known HotL1 elements was present in 35% of patients<sup>17</sup> (Supplementary Table 12). As only one of these events directly affected a known cancer gene (insertion into *ROBO2*), it appears unlikely that this is a major mutational mechanism in PC. No recurrent fusion events were detected (Supplementary Table 13).

## Transcriptional networks and subtypes of PC

We used bulk tumour tissue to better understand the transcriptional networks and molecular mechanisms that underpin the tumour microenvironment. Initial unsupervised clustering of RNA-seq data for 96 tumours with high epithelial content ( $\geq 40\%$ ) to balance stromal gene expression resolved four stable classes (Fig. 1a and Extended Data Fig. 3). These four subtypes were also present in the extended set of 232 PCs using array-based mRNA expression profiles encompassing the full range of tumour cellularity (from 12–100%) (Extended Data Fig. 4). We named these subtypes: (1) squamous; (2) pancreatic progenitor; (3) immunogenic; and (4) aberrantly differentiated endocrine exocrine (ADEX) on the basis of the differential expression of transcription factors and downstream targets important in lineage specification and differentiation during pancreas development and regeneration. Transcriptional network analysis identified 26 coordinately expressed gene programmes representing distinct biological processes, 10 of which discriminated the 4 PC classes (Fig. 1b, Extended Data Fig. 5 and Supplementary Tables 14–16). These 4 subtypes were associated with specific histological characteristics: (1) squamous with adenosquamous carcinomas (6/25 in squamous versus 1/71 in the rest,  $P = 0.0011$  Fisher's exact test); (2) pancreatic progenitor and (3) immunogenic with mucinous non-cystic (colloid) adenocarcinomas and carcinomas arising from IPMN, which are mucinous ( $P = 0.0005$ ); and (4) ADEX with rare acinar cell carcinomas (although numbers were small, both cases clustered with the ADEX class) (Fig. 1a). Squamous subtype was an independent poor prognostic factor (Fig. 1c and Supplementary Table 21).

## Squamous subtype

Four core gene programmes characterized squamous tumours (Fig. 1b), which included gene networks involved in inflammation, hypoxia response, metabolic reprogramming, TGF- $\beta$  signalling, MYC pathway activation, autophagy and upregulated expression of *TP63* $\Delta$ N and its target genes. Many of these genes are highly expressed in the C2-squamous-like class of tumours of breast, bladder, lung and





**Figure 1 | Molecular classes and transcriptional networks defining PDAC.** **a**, Unsupervised analysis of RNA-seq identified 4 PDAC classes: squamous (blue); ADEX (abnormally differentiated endocrine exocrine; brown); pancreatic progenitor (yellow); and immunogenic (red).  $P < 0.05$ , Fisher's exact test. **b**, Heatmap of gene programmes significantly enriched in PDAC. Black dot denotes transcriptional networks showing highest significance for an individual class. **c**, Kaplan–Meier analysis of patient survival stratified by class.

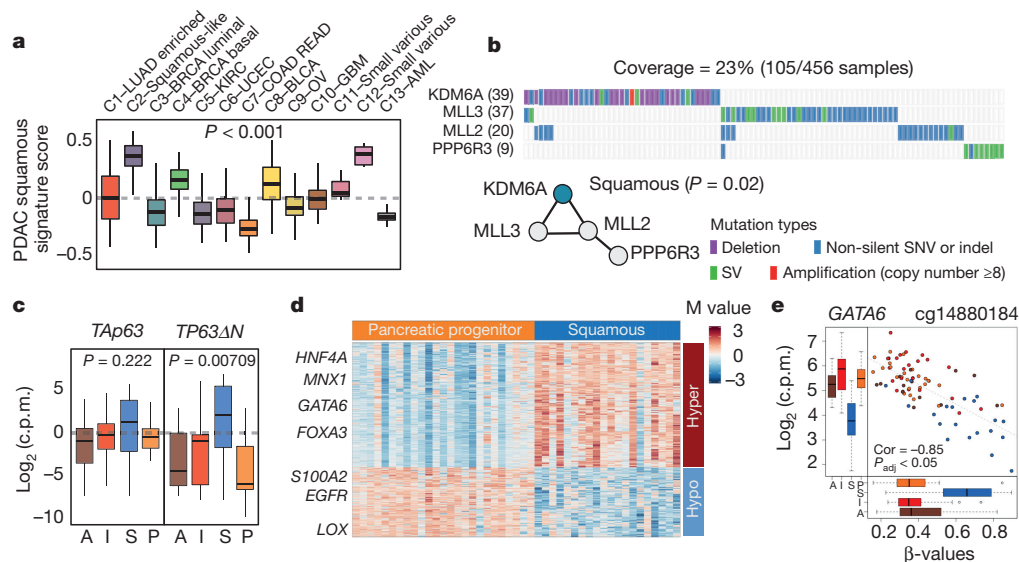
head and neck cancer defined in the Cancer Genome Atlas (TCGA) pan-cancer studies<sup>18</sup>, which was the reason we termed them squamous (Fig. 2a). As in these other cancer types, the pancreatic squamous subtype was associated with mutations in *TP53* ( $P = 0.01$ ) and *KDM6A* ( $P = 0.02$ ), which interacts with ASCOM complex constituents *MLL2* and *MLL3* (Figs 1a and 2b). Although previous immunohistochemical studies have identified increased *TP63* expression in adenosquamous pancreatic tumours<sup>19</sup>, RNA-seq identified high *TP63ΔN* expression and its target genes as a key feature (Fig. 2c). *TP63ΔN*, in the presence of *TP53* mutation, is known to regulate epithelial cell plasticity, tumorigenicity and epithelial to mesenchymal transition in a variety of solid tumours<sup>20</sup>. Squamous tumours were enriched for activated  $\alpha 6\beta 1$  and  $\alpha 6\beta 4$  integrin signalling, and activated EGF signalling. (Extended Data Fig. 6 and Supplementary Table 16). The squamous subtype is associated with hypermethylation and concordant downregulation of genes that govern pancreatic endodermal cell-fate determination (for example, *PDX1*, *MNX1*, *GATA6*, *HNF1B*) leading to a complete loss of endodermal identity (Fig. 2d, e and Supplementary Table 17).

### Pancreatic progenitor subtype

Transcriptional networks containing transcription factors *PDX1*, *MNX1*, *HNF4G*, *HNF4A*, *HNF1B*, *HNF1A*, *FOXA2*, *FOXA3* and *HES1* primarily define the pancreatic progenitor class (Extended Data Fig. 7). These transcription factors are pivotal for pancreatic endoderm cell-fate determination towards a pancreatic lineage and are linked to maturity onset diabetes of the young (MODY). *PDX1*, in particular, is critical for pancreas development with ductal, exocrine and endocrine cells all derived from embryonic progenitor cells that express *PDX1* (ref. 21). Gene programmes regulating fatty acid oxidation, steroid hormone biosynthesis, drug metabolism and O-linked glycosylation of mucins also define pancreatic progenitor tumours. Importantly, apomucins *MUC5AC* and *MUC1*, but not *MUC2* or *MUC6*, are preferentially co-expressed in pancreatic progenitor tumours. The expression of these apomucins defines the pancreatobiliary subtype of IPMN and is consistent with PDAC-associated IPMN clustering within this class (Supplementary Tables 14–16). *TGFBR2* inactivating mutations were also enriched in this subtype ( $P = 0.029$ ).

### ADEX subtype

The ADEX class is defined by transcriptional networks that are important in later stages of pancreatic development and differentiation, and is a subclass of pancreatic progenitor tumours. Transcriptional networks that characterize both exocrine and endocrine lineages at later stages are upregulated, rather than one or the other as is the case in normal pancreas development. The key networks identified include upregulation of: (i) transcription factors *NR5A2*, *MIST1* (also known as *BHLHA15A*) and *RBPJL* and their downstream targets that are important in acinar cell differentiation and pancreatitis/regeneration<sup>22,23</sup>; and (ii) genes associated with endocrine differentiation and MODY (including *INS*, *NEUROD1*, *NKX2-2* and *MAFA* (Extended Data Fig. 8 and Supplementary Table 16)). Importantly, several patient-derived pancreatic cancer cell lines were enriched with gene programmes associated with the ADEX class. Moreover, these cell lines expressed multiple genes associated with terminally differentiated pancreatic tissues, including *AMY2B*, *PRSS1*, *PRSS3*, *CEL* and *INS*. In addition, the methylation pattern of ADEX tumours was distinct from normal pancreas and clustered with other PCs (Extended Data Fig. 9).



**Figure 2 | Molecular characterization of the squamous class.** **a**, Boxplot of PDAC squamous class signature scores generated using pan-cancer 12 expression data and stratified by class. **b**, Mutual exclusivity plot of a mutated gene sub-network identified by HotNet2. **c**, Boxplot of *TAp63* and *TP63ΔN* expression levels stratified by class. **d**, Heatmap of differentially

### Immunogenic subtype

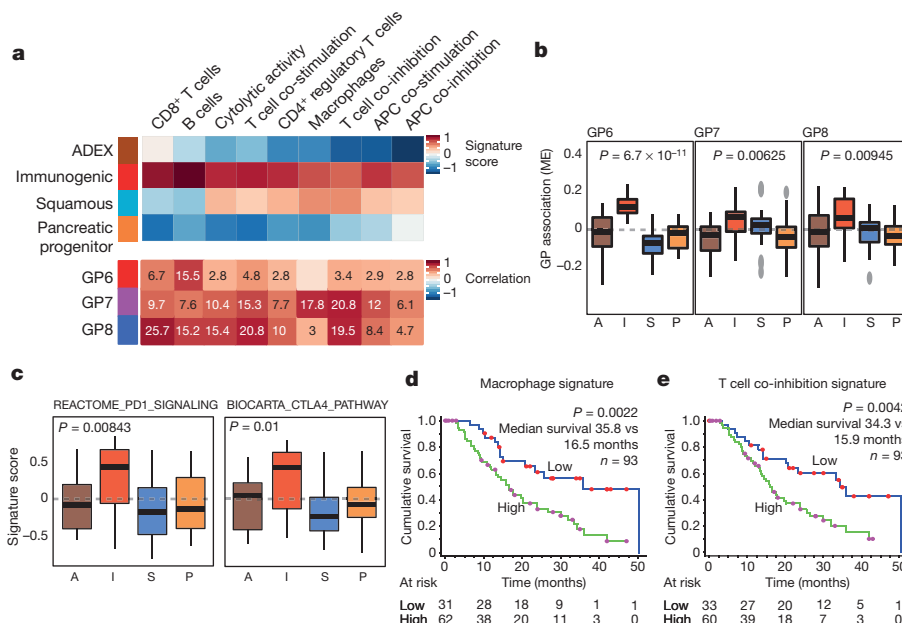
The immunogenic class shares many of the characteristics of the pancreatic progenitor class, but is associated with evidence of a significant immune infiltrate. Associated immune gene programmes included B cell signalling pathways, antigen presentation,  $CD4^+$  T cell,  $CD8^+$  T cell and Toll-like receptor signalling pathways (Extended Data Fig. 10 and Supplementary Table 16). Enrichment analysis identified upregulated expression of genes associated with nine different immune cell types and/or phenotypes<sup>24</sup> (Fig. 3a). The predominant expression profiles were those related to infiltrating B and T cells, with both

methyated genes. **e**, Hypermethylation of *GATA6* is associated with the concordant down regulation of *GATA6* gene expression. Pearson correlation and adjusted *P* values are as indicated. In **a** and **c** the boxplots are annotated by a Kruskal–Wallis *P* value.

cytotoxic ( $CD8^+$ ) and regulatory T cells ( $CD4^+CD25^+FOXP3^+$  T<sub>regs</sub>). Upregulation of CTLA4 and PD1 acquired tumour immune suppression pathways in the immunogenic subtype inferred therapeutic opportunities with novel immune modulators (Fig. 3c).

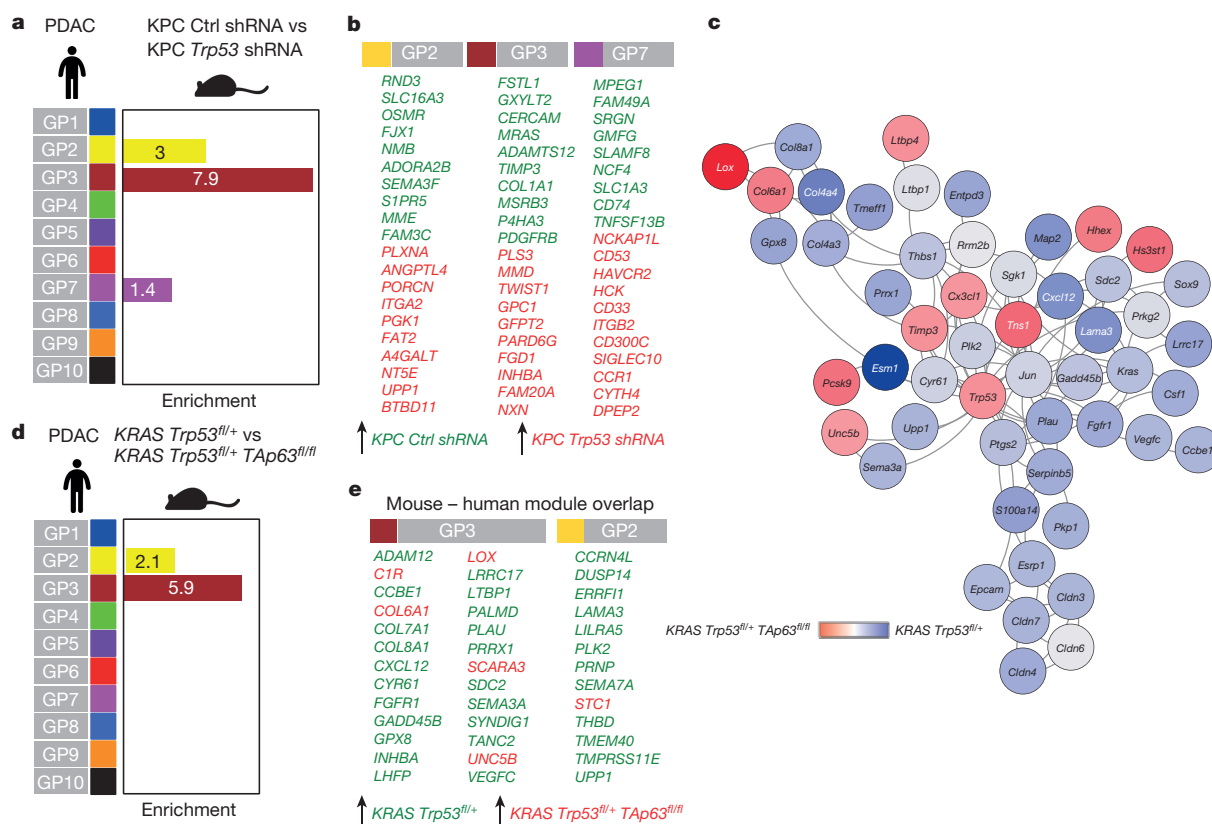
### Immune mechanisms in pancreatic cancer

To better define candidate molecular mechanisms active in the tumour microenvironment, we correlated enrichment of expression patterns that characterize specific immune cell populations with each gene programme (Fig. 3a and Supplementary Tables 15, 16 and 18). Of all gene



**Figure 3 | Immune pathways in PDAC.** **a**, Heatmap showing enrichment of immune cell/phenotype gene signatures in PDAC (top panel). Heatmap showing correlation of immune cell/phenotype gene signatures with the identified PDAC GPs (bottom panel). Numbers in cells represent  $-\log_{10}$  of correlation significance. **b**, Boxplot of GP module eigengene (ME) scores (a measure of sample gene programme relatedness) stratified by class and

showing GP class associations. **c**, Boxplot of *PD1* (also known as *PDCD1*) and *CTLA4* gene signature scores stratified by class. **d**, **e**, Kaplan–Meier analysis comparing survival of patients having either high or low immune cell/phenotype signature scores. In **b** and **c**, the boxplots are annotated by a Kruskal–Wallis *P* value.



**Figure 4 | Gain of function TP53 mutations and loss of TAp63 regulate key GPs associated with the squamous class.** **a**, Significant GP enrichment of genes deregulated in KPC-mouse-derived cell lines treated with Trp53 specific short hairpin RNAs (shRNAs). **b**, Trp53 regulated genes enriched in either GP 2, 3 or 7. **c**, Sub-network of genes differentially expressed between *KRAS Trp53<sup>fl/+</sup>* and *KRAS Trp53<sup>fl/+</sup> Trp63<sup>fl/fl</sup>* cell lines.

Node colour represents change in gene expression. **d**, Genes differentially expressed between *KRAS Trp53<sup>fl/+</sup>* and *KRAS Trp53<sup>fl/+</sup> Trp63<sup>fl/fl</sup>* cell lines significantly enriched in GPs 2 and 3. **e**, Trp63 regulated genes enriched in GPs 2 and 3. In **a** and **d**, bars are annotated with significance values  $-\log_{10}$  ( $P$  value). In **b** and **e**, the arrows and colour represent upregulation of gene expression in the indicated cell types.

programmes (GP), GP6, GP7 and GP8 were enriched with immune cell specific gene expression signatures (Fig. 3b). Specifically, GP6 and GP8 were associated with B cell and CD8<sup>+</sup> T cell signatures, respectively, with GP8 associated with the T cell co-inhibitory phenotype (Extended Data Fig. 10). GP7 was associated with both the macrophage signature and T-cell co-inhibition, which co-segregated with poor survival (Fig. 3d, e). Importantly, pathway analysis of GP7, also showed enrichment for antigen processing and presentation, and Toll-like receptor cascade(s) including high expression of *TLR4*, *TLR7*, *TLR8*, *PDCD1LG2* (*PD-L2*) and *CSF1R*. The latter are known mediators of tumour associated macrophage immunosuppression and inflammation.

### TP53 and TP63 modulation of squamous PDAC

Based on the association of *TP53* mutation and upregulated *TP63* expression in the squamous subtype, we used cell lines derived from genetically engineered mouse models of pancreatic cancer (*Kras<sup>G12D/+</sup>*; *Trp53<sup>fl/+</sup>*; *TAp63<sup>fl/fl</sup>* KPC mice) to begin to unravel the functional consequences of these events in defining squamous tumours. Mice with mutations in the DNA binding domain as compared to *TP53*-null animals have more aggressive disease with increased metastatic potential, primarily mediated through platelet-derived growth factor receptor  $\beta$  (*PDGFRB*)<sup>25</sup>. Analyses of transcriptome data from previous mutant *TP53* knockdown experiments from ref. 25 showed that mutant *TP53* regulates the expression of transcriptional networks associated with the squamous subtype, particularly GPs 2 and 3, including *PDGFRB* (Fig. 4a, b; Supplementary Table 20). *Kras<sup>G12D/+</sup>*; *Trp53<sup>fl/+</sup>*; *TAp63<sup>fl/fl</sup>* mice have more aggressive metastatic pancreatic cancer than their *Kras<sup>G12D/+</sup>*; *Trp53<sup>fl/+</sup>* counterparts and also show deregulation of GPs 2 and 3, inferring that TAp63 plays an important role in squamous PC

(Fig. 4c–e). Transcriptional network analysis identified additional key factors involved in metastasis that were upregulated in the squamous subtype for example, *LOX*<sup>26</sup>.

### Transcriptomic classification of PDAC

We compared our transcriptome classification with those of 2 previously published studies that had either physically<sup>27</sup> or virtually<sup>28</sup> micro-dissected tumour epithelium to define PC subtypes (Fig. 1a and Extended Data Fig. 9). Using their classifiers to subtype our data, 3 of the classes we defined directly overlap with the Collisson classification, with the exception of the novel immunogenic subtype. We altered Collisson's nomenclature to better reflect the insights into the molecular pathology and candidate mechanisms that our integrated analysis generated. The Collisson 'quasimesenchymal' subtype was renamed 'squamous' to reflect the molecular characteristics of squamous tumours across multiple tissue types, as defined by the TCGA pancreatic analysis. 'Classical' was termed 'pancreatic progenitor' based on the prominence of transcriptional networks vital for early pancreas development, and the predominant discriminator from the squamous subtype. The Collisson 'exocrine-like' also contained transcriptional networks characteristic of committed endocrine differentiation and as a consequence were renamed ADEX. Although approximately 50% of squamous subtype tumours fell within the 'basal' subgroup of Moffitt *et al.*<sup>28</sup>, the remainder were composed of a mixture of other Bailey/Collisson subtypes.

More sophisticated analyses using larger numbers of tumours continues to reveal novel insights into pancreatic cancer pathophysiology. In particular, integrated analysis of genomic, epigenomic and transcriptomic characteristics is generating biological insights with



potential therapeutic relevance. The increased appreciation of the role of the immune system in cancer development and progression has led to new classes of therapeutics that specifically target mechanisms through which the tumour evades immune destruction. Therapeutics that target some of these mechanisms are currently in clinical trials in many cancer types, including pancreatic cancer. Early clinical trial data suggest that, similar to most targeted therapies, patient selection will also be important for drugs that target the immune system. The novel immunogenic subtype of pancreatic cancer is characterized by specific mechanisms that can potentially be targeted using immune modulators, and testing in clinical trials is encouraged.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 May; accepted 30 December 2015.**

**Published online 24 February; corrected online 2 March 2016**

(see full-text HTML version for details).

- Rahib, L. *et al.* Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **74**, 2913–2921 (2014).
- Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
- Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806 (2008).
- Wang, L. *et al.* Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by *MLH1* haploinsufficiency and complete deficiency. *Genome Res.* **22**, 208–219 (2012).
- Biakan, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399–405 (2012).
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
- Nones, K. *et al.* Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of *SLIT-ROBO*, *ITGA2* and *MET* signaling. *Int. J. Cancer* **135**, 1110–1118 (2014).
- Nones, K. *et al.* Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nature Commun.* **5**, 5224 (2014).
- Patch, A. M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
- The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
- Bechara, E. G., Sebestyen, E., Bernardis, I., Eyas, E. & Valcarcel, J. RBM5, 6, and 10 differentially regulate *NUMB* alternative splicing to control cancer cell proliferation. *Mol. Cell* **52**, 720–733 (2013).
- Scott, L. M. & Rebel, V. I. Acquired mutations that affect pre-mRNA splicing in hematologic malignancies and solid tumors. *J. Natl. Cancer Inst.* **105**, 1540–1549 (2013).
- Maguire, S. L. *et al.* *SF3B1* mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* **235**, 571–580 (2015).
- Horn, S. *et al.* *Mind bomb 1* is required for pancreatic  $\beta$ -cell formation. *Proc. Natl Acad. Sci. USA* **109**, 7356–7361 (2012).
- Scaltriti, M. *et al.* Cyclin E amplification/overexpression is a mechanism of trastuzumab resistance in HER2<sup>+</sup> breast cancer patients. *Proc. Natl Acad. Sci. USA* **108**, 3761–3766 (2011).
- Shain, A. H., Salari, K., Giacomini, C. P. & Pollack, J. R. Integrative genomic and functional profiling of the pancreatic cancer genome. *BMC Genomics* **14**, 624 (2013).
- Tubio, J. M. *et al.* Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**, 1251343 (2014).
- Hoadley, K. A. *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929–944 (2014).
- Brody, J. R. *et al.* Adenosquamous carcinoma of the pancreas harbors *KRAS2*, *DPC4* and *TP53* molecular alterations similar to pancreatic ductal adenocarcinoma. *Mod. Pathol.* **22**, 651–659 (2009).
- Engelmann, D. & Putzer, B. M. Emerging from the shade of p53 mutants: N-terminally truncated variants of the p53 family in EMT signaling and cancer progression. *Sci. Signal.* **7**, re9 (2014).
- Hale, M. A. *et al.* The homeodomain protein *PDX1* is required at mid-pancreatic development for the formation of the exocrine pancreas. *Dev. Biol.* **286**, 225–237 (2005).

- von Figura, G., Morris, J. P. IV, Wright, C. V. & Hebrok, M. Nr5a2 maintains acinar cell differentiation and constrains oncogenic Kras-mediated pancreatic neoplastic initiation. *Gut* **63**, 656–664 (2014).
- Hale, M. A. *et al.* The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis. *Development* **141**, 3123–3133 (2014).
- Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
- Weissmueller, S. *et al.* Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor  $\beta$  signaling. *Cell* **157**, 382–394 (2014).
- Miller, B. W. *et al.* Targeting the LOX/hypoxia axis reverses many of the features that make pancreatic cancer deadly: inhibition of LOX abrogates metastasis and enhances drug efficacy. *EMBO Mol. Med.* **7**, 1063–1076 (2015).
- Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature Med.* **17**, 500–503 (2011).
- Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nature Genet.* **47**, 1168–1178 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We would like to thank C. Axford, M.-A. Brancato, S. Rowe, M. Thomas, S. Simpson and G. Hammond for central coordination of the Australian Pancreatic Cancer Genome Initiative, data management and quality control; M. Martyn-Smith, L. Braatvedt, H. Tang, V. Papangelis and M. Beilin for biospecimen acquisition; and Deborah Gwynne for support at the Queensland Centre for Medical Genomics. We also thank M. Hodgins, M. Debeljak and D. Trusty for technical assistance at Johns Hopkins University. Funding support was from: National Health and Medical Research Council of Australia (NHMRC; 631701, 535903, 427601); Queensland Government (NIRAP); University of Queensland; Australian Government: Department of Innovation, Industry, Science and Research (DIISR); Australian Cancer Research Foundation (ACRF); Cancer Council NSW (SRP06-01, SRP11-01, ICGC); Cancer Institute NSW (10/ECF/2-26; 06/ECF/1-24; 09/CDF/2-40; 07/CDF/1-03; 10/CRF/1-01, 08/RSA/1-15, 07/CDF/1-28, 10/CDF/2-26, 10/FRL/2-03, 06/RSA/1-05, 09/RIG/1-02, 10/TPG/1-04, 11/REG/1-10, 11/CDF/3-26); Garvan Institute of Medical Research; Cancer Research UK Glasgow Centre Program, A18076; Avner Nahmani Pancreatic Cancer Research Foundation; R.T. Hall Trust; Petre Foundation; Philip Hemstritch Foundation; Gastroenterological Society of Australia (GESA); American Association for Cancer Research (AACR) Landon Foundation—INNOVATOR Award; Wellcome Trust Senior Investigator Award 103721/Z/14/Z; Cancer Research UK Programme Grant C29717/A17263; Cancer Research UK Programme Grant A12481; Pancreatic Cancer UK; The Howat Foundation; University of Glasgow; European Research Council Starting Grant, 311301, Italian Ministry of University and Research (Cancer Genome Project FIRB RBAP10AHJB), Associazione Italiana Ricerca Cancro (n.12182), Fondazione Italiana Malattie Pancreas – Ministry of Health (CUP\_J33G13000210001), European Community Grant FP7 Cam-Pac, grant agreement number 602783.

**Author Contributions** Investigator contributions are as follows: P.J.B., J.V.P., N.W., A.V.B., S.M.G. (concept and design); P.J.B., D.A.W., R.A.G., A.S., D.K.C., J.V.P., N.W., A.V.B., S.M.G. (project leaders); P.J.B., D.K.C., A.V.B., S.M.G. (writing team); D.K.M., A.N.C., T.J.C.B., C.N., K.N., S.W., D.M.M., N.W., L.E., L.M., L.S., S.M.G., I.H., S.I., S.M., E.N., K.Q., S.M.G. (genomics); P.J.B., D.K.M., K.S.K., N.W., P.J.W., O. H., A.M.P., F.N., O.H., C.L., D.T., S.W., Q.X., K.N., N.C., M.Q., M.A., A.R., M.G., S.K., K.Q., L.P., J.M., M.C., L.C.M., O.S., L.F., U.B., N.W., J.V.P., S.M.G. (data analysis); D.K.C., A.L.J., A.M.N., A.M., A.V.P., C.W.T., E.K.C., E.S.H., I.R., M.G., J.H., J.A.L., K.E., L.A.C., M.D.J., A.J.G., N.Q.N., A.B., N.Z., C.P., R.G., J.R.E., R.H.H., A.M., C.A.I., C.L.W., B.R., V.C., P.C., C.B., R.S., G.T., D.M., G.M.P., J.H., M.P., J.W., V.C., C.J.S., J.G.K., R.T.L., N.D.M., N.B.J., J.S.G., J.D.S., R.A.M., J.H., S.A.K., K.M., R.L.S., A.V.B. (sample acquisition and processing, clinical annotation, interpretation and analysis); A.J.G., A.C., R.H.H., F.D., K.O., A.S., W.F., J.G.K., C.T. (pathology assessment).

**Author Information** All DNA sequencing and RNA-seq data have been deposited in the European Genome-phenome Archive (EGA); accession code EGAS00001000154. All gene expression, genotyping, and methylome data used in this study has been deposited in the NCBI Gene Expression Omnibus (GEO) under accession codes GSE49149 and GSE36924. Mouse cell line expression data are available in the ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress>) under accession number E-MTAB-4415. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.V.B. (andrew.biakan@glasgow.ac.uk) or S.M.G. (sean.grimmond@unimelb.edu.au).

## METHODS

**Human research ethical approvals.** APGI: Sydney South West Area Health Service Human Research Ethics Committee, western zone (protocol number 2006/54); Sydney Local Health District Human Research Ethics Committee (X11-0220); Northern Sydney Central Coast Health Harbour Human Research Ethics Committee (0612-251M); Royal Adelaide Hospital Human Research Ethics Committee (091107a); Metro South Human Research Ethics Committee (09/QPAH/220); South Metropolitan Area Health Service Human Research Ethics Committee (09/324); Southern Adelaide Health Service/Flinders University Human Research Ethics Committee (167/10); Sydney West Area Health Service Human Research Ethics Committee (Westmead campus) (HREC2002/3/4.19); The University of Queensland Medical Research Ethics Committee (2009000745); Greenslopes Private Hospital Ethics Committee (09/34); North Shore Private Hospital Ethics Committee. Johns Hopkins Medical Institutions: Johns Hopkins Medicine Institutional Review Board (NA00026689). ARC-Net, University of Verona: approval number 1885 from the Integrated University Hospital Trust (AOUI) Ethics Committee (Comitato Etico Azienda Ospedaliera Universitaria Integrata) approved in their meeting of 17 November 2010 and documented by the ethics committee 52070/CE on 22 November 2010 and formalized by the Health Director of the AOUI on the order of the General Manager with protocol 52438 on 23 November 2010. Ethikkommission an der Technischen Universität Dresden (Approval numbers EK30412207 and EK357112012).

**Patient material acquisition and extraction.** Samples were acquired through the Australian Pancreatic Cancer Genome Initiative (APGI) as part of the International Cancer Genome Consortium (ICGC). Informed consent was obtained from all subjects. Tissue dissection of primary material, RNA and DNA extraction was performed using previously published methods<sup>2</sup>. Tumour cellularity was estimated for each sample using a combination of qPure analysis of high-density SNP profiles and *KRAS* amplicon sequencing<sup>2</sup>. Primary tumours ( $n=342$ ) and 41 patient-derived cell lines (representing low cellularity tumours) (Supplementary Table 1) underwent whole genome sequencing when tumour cellularity was  $>40\%$  (mean coverage  $75\times$ ,  $n=179$ ), or deep-exome sequencing (mean coverage:  $400\times$ ,  $n=204$ ) for samples with a cellularity of 12–40%.

**Exome library preparation.** Exome libraries were generated using the Illumina Nextera Rapid Capture Exome kit (Illumina, Part no. FC-140–1003) according to the standard manufacturer's protocol (part no. 15037436 Rev. A February 2013), except they were made in an automated high-throughput fashion using Perkin Elmer's Sciclone G3 NGS Workstation (Product no. SG3-31020-0300). Then 50 ng of gDNA was used as input for fragmentation followed by 10 cycles of PCR to produce sufficient library for exome capture. A total of 500 ng of each library was pooled as a 12-plex reaction for capture using Illumina's Nextera Exome Oligo set. Following two rounds of capture, samples were finally subjected to 10 cycles of PCR to produce exome libraries ready for sequencing. Prior to sequencing, exome libraries were qualified via either the Perkin Elmer LabChip GX with the DNA High Sensitivity LabChip kit (Perkin Elmer, Part no. CLS760672), or the Agilent BioAnalyzer 2100 with the High Sensitivity DNA Kit (Agilent, Part no. 5067–4626). Quantification of libraries for clustering was performed using the KAPA Library Quantification Kit - Illumina/Universal (KAPA Biosystems, Part no. KK4824) in combination with the Life Technologies Viia 7 real time PCR instrument.

**Whole-genome library preparation.** Whole-genome libraries were generated using either the Illumina TruSeq DNA LT sample preparation kit (Illumina, Part no. FC-121–2001 and FC-121–2001) or the Illumina TruSeq DNA PCR-free LT sample preparation kit (Illumina, Part no. FC-121–3001 and FC-121–3002) according to the manufacturer's protocols with some modifications (Illumina, Part no. 15026486 Rev. C July 2012 and 15036187 Rev. A January 2013 for the two different kits respectively). For the TruSeq DNA LT sample preparation kit, 1  $\mu$ g of gDNA was used as input for fragmentation to  $\sim 300$  bp, followed by a SPRI-bead clean up using the AxyPrep Mag PCR Clean-Up kit (Corning, Part no. MAG-PCR-CL-250). After end-repair, 3' adenylation and adaptor ligation, the libraries were size-selected using a double SPRI-bead method to obtain libraries with an insert size  $\sim 300$  bp. The size-selected libraries were subjected to 8 cycles of PCR to produce the final whole-genome libraries ready for sequencing. For the TruSeq DNA PCR-free LT sample preparation kit, 1  $\mu$ g of gDNA was used as input for fragmentation to  $\sim 350$  bp, followed by an end-repair step and then a size-selection using the double SPRI-bead method to obtain libraries with an insert size  $\sim 350$  bp. The size-selected libraries then underwent 3' adenylation and adaptor ligation to produce final whole genome libraries ready for sequencing. Prior to sequencing, whole-genome libraries were qualified via the Agilent BioAnalyzer 2100 with the High Sensitivity DNA Kit (Agilent, Part no. 5067–4626). Quantification of libraries for clustering was performed using the KAPA Library Quantification Kit - Illumina/Universal (KAPA Biosystems, Part no. KK4824) in combination with the Life Technologies Viia 7 real time PCR instrument.

**Total RNA library preparation.** RNA-Seq libraries were generated using the Illumina TruSeq Stranded Total RNA LT sample preparation kit (with Ribo-Zero Gold) (Illumina, Part no. RS-122–2301 and RS-122–2302), according to the standard manufacturer's protocol (Part no. 15031048 Rev. D April 2013), except they were made in an automated high-throughput fashion using Perkin Elmer's Sciclone G3 NGS Workstation (Product no. SG3-31020-0300). The ribosomal depletion step was performed on 1  $\mu$ g of total RNA using Ribo-Zero Gold before a heat fragmentation step aimed at producing libraries with an insert size between 120–200 bp. cDNA was then synthesized from the enriched and fragmented RNA using SuperScript II Reverse Transcriptase (Invitrogen, Catalog no. 18064) and random primers. The resulting cDNA was converted into double-stranded DNA in the presence of dUTP to prevent subsequent amplification of the second strand and thus maintain the strandedness of the library. Following 3' adenylation and adaptor ligation, libraries were subjected to 15 cycles of PCR to produce RNA-seq libraries ready for sequencing. Prior to sequencing, RNA-seq libraries were qualified via the Perkin Elmer LabChip GX with the DNA High Sensitivity LabChip kit (Perkin Elmer, Part no. CLS760672). Quantification of libraries for clustering was performed using the KAPA Library Quantification Kit - Illumina/Universal (KAPA Biosystems, Part no. KK4824) in combination with the Life Technologies Viia 7 real time PCR instrument.

**Library sequencing.** All libraries were sequenced using the Illumina HiSeq 2000/2500 system with TruSeq SBS Kit v3 - HS (200-cycles) reagents (Illumina, Part no. FC-401-3001), to generate paired-end 101 bp reads.

**Sequence alignment and data management.** Sequence data was mapped to the Genome Reference Consortium GRCh37 assembly using BWA42. All BAM files have been deposited in the EGA (accession number: EGAS00001000154).

**Copy number analysis.** Matched tumour and normal patient DNA was assayed using Illumina SNP BeadChips as per manufacturer's instructions (Illumina, San Diego CA) (HumanOmni1-Quad or HumanOmni2.5–8 BeadChips) and analysed as previously described<sup>2,8</sup>.

**Identification and verification of structural variants.** The Somatic structural variant pipeline were identified using the qSV tool. A detailed description of its use has been recently published<sup>2,8</sup>.

**Identification of and verification of point mutations.** Substitutions and indels were called using a consensus calling approach that included qSNP, GATK and Pindel. The details of call integration and filtering, and verification using orthogonal sequencing and matched sample approaches are as previously described<sup>2,8,9</sup>. 97% of *KRAS* mutations identified by *KRAS* deep-amplicon sequencing were detected via WGS and WES, inferring a false negative rate of 3% (Supplementary Table 1).

**'Lollipop' plots.** Plots showing the location and frequency of inactivating mutations were generated using the MutationMapper web tool hosted at <http://www.cbioportal.org/>. Available PanCancer mutation data was downloaded from the Cancer Genomic Data Server (CGDS) hosted by the Computational Biology Center (cBio) at the Memorial Sloan-Kettering Cancer Center (MSKCC) using the R package "cgdsr"<sup>29</sup>.

**Mutational signatures.** Mutational signatures were defined for genome-wide somatic substitutions, as previously described<sup>2</sup>.

**Significantly mutated gene detection.** A combination of three robust approaches were used to define significantly mutated genes: (i) MutSigCV2 (ref. 30), which detects genes with point mutations above the background mutation rate; (ii) OncodriverFM<sup>31</sup>, which detects point mutated genes with a bias towards pathogenic mutations; and (iii) HOTNET2 (ref. 32), which identifies sub-networks based on protein–protein interactions that contain recurrent point mutations, copy number alterations and structural rearrangements. The HotNet2 (HotNet diffusion-oriented subnetworks) algorithm was used to identify significantly mutated subnetworks in a genome-scale interaction network. Heat scores for each protein were calculated as the number of samples having a non-silent SNV, indel, SV or copy number aberration in the corresponding gene<sup>32</sup>. Heat scores were limited to proteins having a corresponding gene mutation in  $\geq 2\%$  of samples. The iRefIndex interaction network was used for the analysis<sup>33</sup>. Supplementary Table 20 contains matrices summarizing all mutations, CNVs and SVs for all samples used in this study.

**RNA sequencing library generation and sequencing.** RNA-seq libraries were generated using TruSeq Stranded Total RNA (part no. 15031048 Rev. D April 2013) kits, using on a Perkin Elmer's Sciclone G3 NGS Workstation (product no. SG3-31020-0300). Ribosomal depletion step was performed on 1  $\mu$ g of total RNA using Ribo-Zero Gold before a heat fragmentation step aimed at producing libraries with an insert size between 120–200 bp. cDNA was then synthesized from the enriched and fragmented RNA using Invitrogen's SuperScript II Reverse Transcriptase (catalogue number 18064) and random primers. The resulting cDNA was further converted into double stranded DNA in the presence of dUTP to prevent



subsequent amplification of the second strand and thus maintain the strandedness of the library. Following 3' adenylation and adaptor ligation libraries were subjected to 15 cycles of PCR to produce RNA-seq libraries ready for sequencing. Prior to sequencing, exome and RNA-seq libraries were qualified and quantified via Caliper's LabChip GX (part no. 122000) instrument using the DNA High Sensitivity Reagent kit (product no. CLS760672). Quantification of libraries for clustering was performed using the KAPA Library Quantification Kits For Illumina sequencing platforms (kit code KK4824) in combination with Life Technologies Viia 7 real time PCR instrument.

**RNA-seq analysis.** Sequencing reads were mapped to transcripts corresponding to ensemble 70 annotations using RSEM<sup>34</sup>. RSEM data were normalized using TMM (weighted trimmed mean of M-values) as implemented in the R package 'edgeR'. For downstream analyses, normalized RSEM data were converted to counts per million (c.p.m.) and  $\log_2$  transformed<sup>35</sup>. Genes without at least 1 c.p.m. in 20% of the sample were excluded from further analysis.

**RNA-seq re-analysis of Weismuller *et al.*** RNA sequencing data reported in ref. 25 was downloaded from the Sequence Read Archive (SRA): Accession number: SRP033333. The available data was re-analysed using an RNA-seq pipeline implemented in the bcbio-nextgen project (<https://bcbio-nextgen.readthedocs.org/en/latest/>). Briefly, after quality control and adaptor trimming, reads were aligned to the UCSC mouse mm10 genome build using STAR<sup>36</sup> Counts for known genes were generated using the function `featureCounts` in the R/Bioconductor package 'Rsubread'<sup>37</sup>. The R/Bioconductor package 'DESeq2' was used to identify differentially expressed genes<sup>38</sup>.

**KRAS *Trp53*<sup>fl/+</sup> and KRAS *Trp53*<sup>fl/+</sup> *Trp63*<sup>fl/fl</sup> mouse derived cell lines.** Cell lines were generated in house from pancreatic tumours harvested from *Pdx1-Cre, LSL-Kras*<sup>G12D/+</sup>, *Trp53*<sup>fl/+</sup> mice or *Pdx1-Cre, LSL-Kras*<sup>G12D/+</sup>, *Trp53*<sup>fl/+</sup>, *Trp63*<sup>fl/fl</sup> mice described previously<sup>49</sup>. Low passage cell lines were used and authenticated by morphology. Mycoplasma testing confirmed that all cell lines were mycoplasma negative. Independently derived cell-lines representing either the *KRAS Trp53*<sup>fl/+</sup> ( $n=3$ ) or *KRAS Trp53*<sup>fl/+</sup> *Trp63*<sup>fl/fl</sup> ( $n=3$ ) genotype were used for RNA-seq analysis. RNA-seq libraries were generated using the KAPA stranded RNaseq Kit with RiboErase (HMR) (KAPA Biosystems; kit ref. KR1151 – v3.15) according to the manufacturer's instructions. Briefly, samples were fragmented for 6 min at 94 °C with 10 cycles of library amplification. Library quality control was performed using an Agilent BioAnalyzer 2100 in combination with a High Sensitivity DNA Kit (Agilent, Part no. 5067-4626). Samples were evenly pooled to a 2 nM concentration and a 1% PhiX control spike-in was used for sequencing quality control. Libraries were run on the NextSeq 500 platform according to the manufacturer's instructions (Illumina, San Diego CA). Sequenced libraries were mapped to UCSC mouse mm10 genome build using TopHat and differential gene expression determined using Cufflinks 2.1.1 and Cuffdiff 2.1.1 as implemented in BaseSpace (<https://basespace.illumina.com/home/indexIllumina>, San Diego CA).

**Microarray analysis.** Tumour RNA was assayed using HumanHT-12 v4 Expression BeadChips as per manufacturer's instructions (Illumina, San Diego CA) and analysed as previously described<sup>7</sup>. Batch correction was performed using the R package 'sva'<sup>39</sup>.

**Clustering.** Non-negative matrix factorization (NMF) was employed to identify stable sample clusters<sup>40</sup>. The top 2,000 most variable genes were used as input. NMF parameters: Brunet algorithm;  $k=1$  to  $k=7$  clusters; number of clusterings to build consensus matrix = 20; error function = Euclidean; and 500 iterations. The preferred clustering result was determined using the observed cophenetic correlation between clusters and the average silhouette width of the consensus membership matrix as determined by the R package 'cluster'. The R package 'ConsensusClusterPlus'<sup>41</sup> was also employed to verify sample clustering. Similar sample clusters were obtained using both methods (data not shown). The package 'ConsensusClusterPlus' was also used to subtype PC samples according to the expression signatures defined in Moffitt *et al.*<sup>28</sup>

**Differential gene expression (DGE).** To identify the most representative samples within each cluster, we computed silhouette widths using the R 'cluster' package. Samples with positive silhouette widths were retained for DGE analysis. DGE analysis between representative samples was performed using the function 'voom' as implemented in the R package 'edgeR'<sup>42</sup>. To define genes differentially expressed between all classes we used the function 'sam' as implemented in the R package 'siggenes'.

**Gene sets.** Gene sets representing immune cell-type expression markers and immune meta-genes were obtained from a recent publication<sup>24</sup>. Gene sets representing PDAC classes were generated by selecting significantly upregulated genes in a given class versus all other classes. An adjusted  $P$  value of 0.01 was used as the cut-off in each case.

**Gene set enrichment.** Gene set enrichment was performed using the R package 'GSVA' (function `gsva` - arguments: `method = "gsva"`, `mx.diff = TRUE`)<sup>43</sup>. GSVA implements a non-parametric unsupervised method of gene set enrichment

that allows an assessment of the relative enrichment of a selected pathway across the sample space. The output of GSVA is a gene-set by sample matrix of GSVA enrichment scores that are approximately normally distributed. GSVA enrichment scores were generated for each gene set using the transformed RSEM data unless otherwise indicated. For survival analyses, sample GSVA enrichment scores were stratified into quantiles (for example, lower 33% or upper 66% of values).

**WGCNA.** Weighted gene co-expression network analysis (WGCNA) was used to generate a transcriptional network from the normalized and transformed RSEM<sup>44</sup>. Briefly, WGCNA clusters genes into network modules using a topological overlap measure (TOM). The TOM is a highly robust measure of network interconnectiveness and essentially provides a measure of the connection strength between two adjacent genes and all other genes in a network. Genes are clustered using 1-TOM as the distance measure and gene modules are defined as branches of the resulting cluster tree using a dynamic branch-cutting algorithm<sup>45</sup>.

The module eigengene is used as a measure of module expression in a given sample and is defined as the first principle component of a module. To relate sample traits of interest to gene modules, sample traits were correlated to module eigengenes and significance determined by a Student asymptotic  $P$  value for the given correlations. For gene module survival analyses, module eigengenes were stratified into quantiles (for example, lower 33% or upper 66% of values). To relate gene modules to PDAC classes, PDAC class gene set GSVA enrichment scores were used as sample traits and correlated with the module eigengenes as discussed above. Similarly, to relate the immune cell-type expression markers and immune meta-genes to the gene modules each immune GSVA enrichment score was correlated with the module eigengenes as before.

To determine the enrichment of differentially expressed mouse genes in modules generated by WGCNA, mouse identifiers were first mapped to their corresponding human HGCN Symbol using the R/Bioconductor package 'biomaRt'. Module gene enrichment was then determined using the function `userListEnrichment` in the WGCNA package. We considered, as significant, only those modules showing both significant enrichment and significant gene expression/gene module correlations.

**Pathway analysis.** Ontology and pathway enrichment analysis was performed using the R package 'dnet'<sup>46</sup> and/or the Reactome FI Cytoscape plugin 4.1.1 (ref. 47) as indicated. The R package 'dnet' was also used to identify significant sub-networks of differentially expressed genes.

**Pan-cancer 12 data and squamous assignment.** Platform corrected input data was obtained from Synapse as part of the Pan-Cancer 12 data freeze (syn1715755)<sup>18</sup>. Pan-cancer 12 subtype assignments were also obtained from Synapse (syn1889916) and sample sizes, as indicated, used for statistical comparisons. To determine the relationship between the PDAC classes and the pan-cancer 12 subtypes, PDAC class gene sets were used in combination with the pan-cancer 12 expression data to generate GSVA enrichment scores as discussed above. Sample GSVA enrichment scores representing each PDAC class were then stratified according to the pan-cancer 12-subtype assignments. A Kruskal-Wallis test was applied to the stratified scores to determine whether the distributions differed.

**Methylation analysis.** Sample methylation was determined using Illumina 450K arrays as previously described<sup>7</sup> with the following modifications. Probe-level Illumina GeneStudio output files were imported into R package 'lumi'<sup>48</sup> and data filtered to remove failed hybridizations, probes comprising SNPs and probes located on sex chromosomes. The filtered methylation values were then colour balance corrected and normalized using Shift and Scaling Normalization (SSN) as implemented by lumi. Gene methylation values were obtained by collapsing probe level values for a given gene loci (that is, probes located 1,500 bp upstream of the transcriptional start site (TSS) through to the end of transcription) using the function `collapseRows(method = "maxRowVariance")` from the package WGCNA<sup>44</sup>. Probes were mapped to gene loci using the R package 'genomicFeatures'. Differential gene methylation between representative samples (selected as above under heading Differential gene expression (DGE)) was determined using the R package 'limma'.  $M$ -values were used for differential gene methylation analysis. Concordant changes in methylation and expression were calculated as follows. Probes were mapped to a given gene using the R package 'genomicFeatures'. As above, probes located 1,500 bp upstream of the TSS through to the end of transcription were considered for each gene. The correlation between a probe  $\beta$ -value and the corresponding gene  $\log_2$  (CPM) expression value was then calculated using Pearson's correlation coefficient. The statistical significance of each probe/gene correlation was calculated by permuting the data 10,000 times and comparing the correlation coefficients obtained before and after permutation. The resulting  $P$  values were adjusted for multiple testing using the approach of Benjamini and Hochberg.

**Concordant copy number expression analysis.** Analysis of variance was used to identify significant changes in gene expression between samples exhibiting corresponding gene copy number aberrations. Accordingly, gene expression



values were stratified on the basis of sample copy number change for a given gene. Deleted genes and genes having copy number  $\geq 8$  were considered for the analysis. *P* values were adjusted using the R package 'qvalue' for which adjusted values  $< 0.05$  were considered statistically significant. Variance was similar between groups compared. Genes showing copy number aberrations in less than 3% of samples were excluded from the analysis.

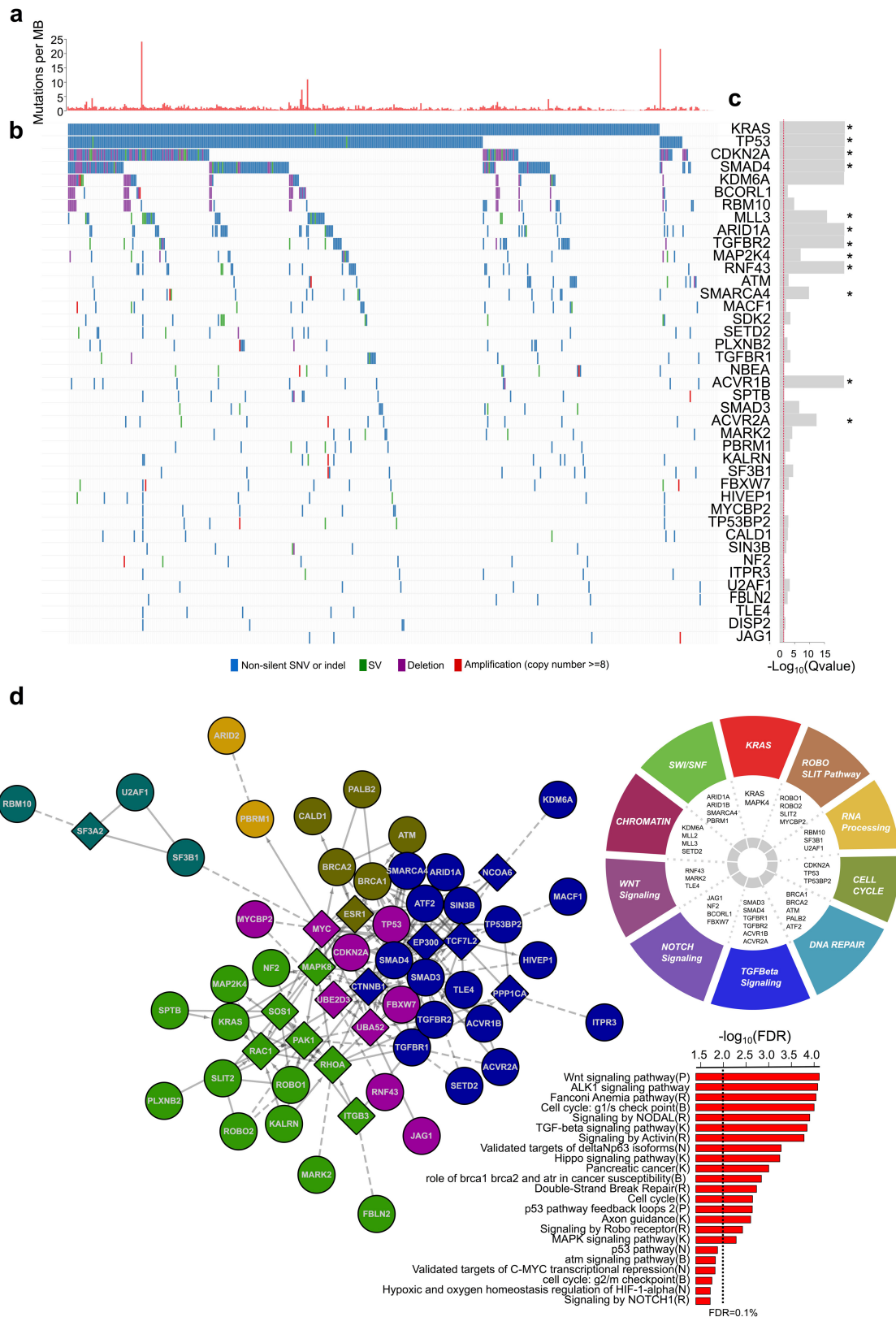
**Survival analysis.** The date of diagnosis and the date and cause of death were obtained from the Central Cancer Registry and treating clinicians. Median survival was estimated using the Kaplan–Meier method and the difference was tested using the log-rank test. *P* values of less than 0.05 were considered statistically significant. Clinicopathologic variables analysed with a *P* value  $< 0.25$  on log-rank test were entered into Cox proportional hazards multivariate analysis. Statistical analysis was performed using StatView 5.0 Software (Abacus Systems, Berkeley, CA, USA). Disease-specific survival was used as the primary endpoint.

**Stromal cell and immune infiltrate quantification.** To quantify stromal and immune cell tumour contributions we used the R package 'estimate'<sup>50</sup>.

**Statistical analysis.** A Kruskal–Wallis test was applied to the indicated stratified scores to determine whether distributions were significantly different. For all PDAC class comparisons using RNA-seq data, the following sample sizes were compared: ADEX ( $n = 14$ ), immunogenic ( $n = 24$ ); squamous ( $n = 20$ ); and pancreatic progenitor ( $n = 25$ ). Fisher's exact tests were used to evaluate the association between dichotomous variables.

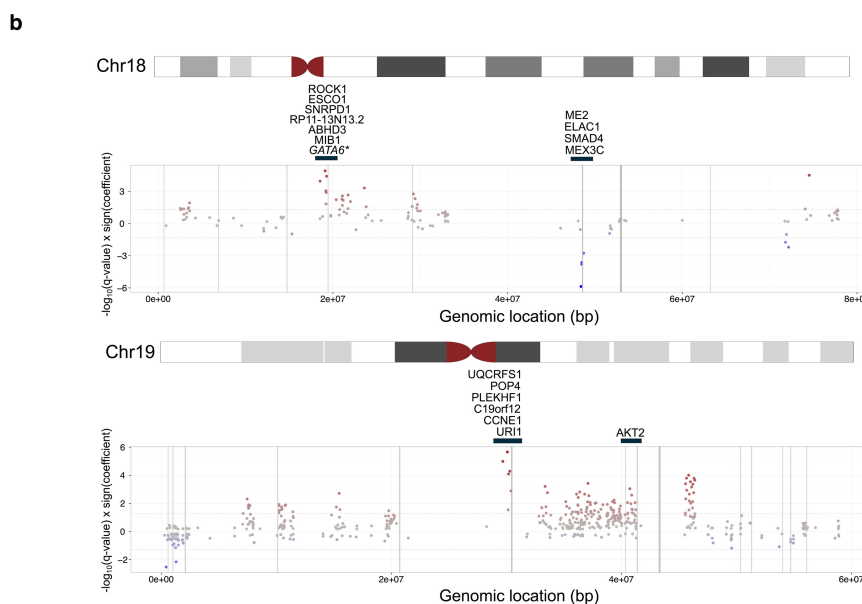
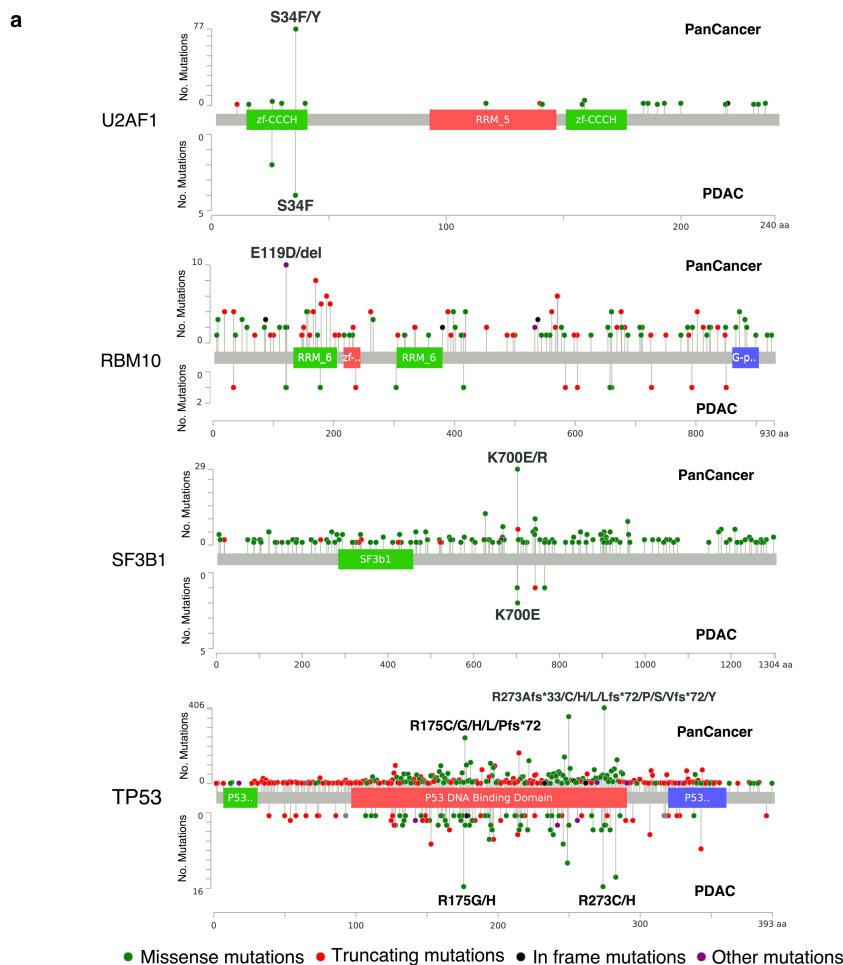
**Data.** Data presented in this study can be downloaded from the following repository <https://dcc.icgc.org/repositories> under the identifier PACA-AU.

29. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
30. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
31. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nature Methods* **10**, 1081–1082 (2013).
32. Leiserson, M. D. *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature Genet.* **47**, 106–114 (2015).
33. Razick, S., Magklaras, G. & Donaldson, I. M. iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics* **9**, 405 (2008).
34. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
35. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
36. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
37. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
38. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
39. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
40. Kuehn, H., Liberzon, A., Reich, M. & Mesirov, J. P. Using GenePattern for gene expression analysis. *Current Protoc. Bioinformatics* Chapter 7, Unit 7 12, (2008).
41. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
42. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
43. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
44. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
45. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
46. Fang, H. & Gough, J. The 'dnet' approach promotes emerging research on cancer patient survival. *Genome Med.* **6**, 64 (2014).
47. Wu, G. & Stein, L. A network module-based method for identifying cancer prognostic signatures. *Genome Biol.* **13**, R112 (2012).
48. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
49. Tan, E. H. *et al.* Functions of Tap63 and p53 in restraining the development of metastatic cancer. *Oncogene* **25**, 3325–3333 (2014).
50. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Commun.* **4**, 2612 (2013).



**Extended Data Figure 1 | Mutational landscape of PC. a**, Barplot representing the somatic mutation rate for each of the 456 samples included in this analysis. **b**, Non-silent mutations (blue), amplifications ( $\geq 8$  copies, red), deletions (purple) and structural variants (SV, green) ranked in order of exclusivity. **c**, Significantly mutated genes identified by OncodriverFM. An asterisk denotes a significantly mutated gene identified by both MutSigCV and OncodriverFM. **d**, PC mutation functional interaction (FI) sub-network identified by the ReactomeFI cytoscape

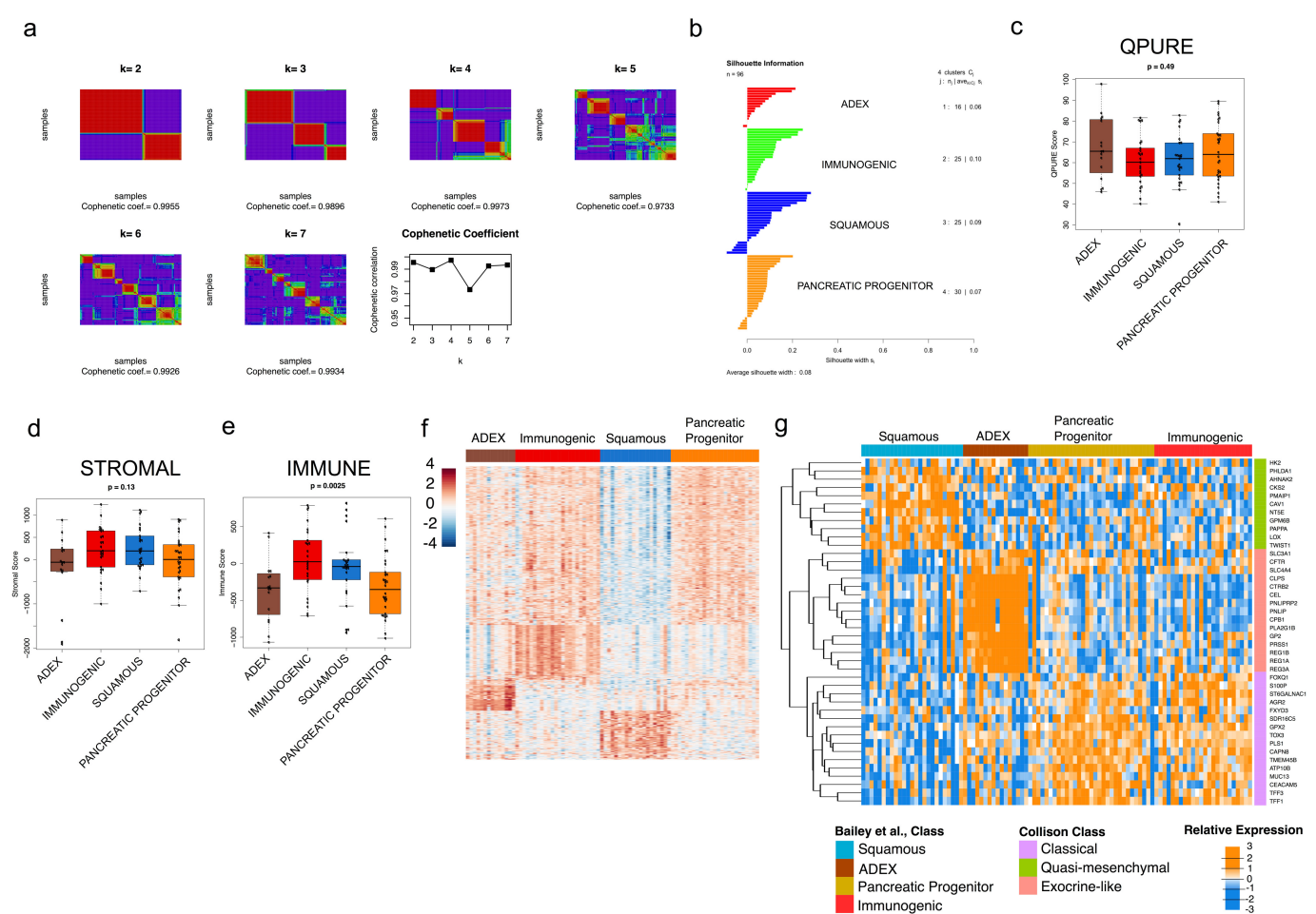
plugin. Mutated genes are indicated as coloured circles and linker genes (that is, genes not significantly mutated but highly connected to mutated genes in the network) indicated as coloured diamonds. Different node colours indicate different network clusters or closely interconnected genes.  $P$  values represent  $FDR < 0.05$ . Pathways significantly enriched in the identified FI sub-network are shown in the accompanying bar graph. Linker genes were not included in the enrichment analysis. Pie chart representing significantly altered genes and pathways in PC.



**Extended Data Figure 2 | Selected genomic events in PC.** **a**, Lollipop plots showing the type and location of mutations in the RNA processing genes *RBM10*, *SF3B1* and *U2AF1* and the tumour suppressor *TP53*. In each plot, mutations observed across multiple cancers (top plot; PanCancer) are compared with those observed in the current study (bottom plot; PDAC). Significant recurrent mutations are labelled above the relevant lollipop. **b**, Regions of copy number alteration showing concordant gene expression changes. For each of the indicated chromosomes, significant GISTIC peaks are shown at their respective genomic locations (x axis) as grey bars. Each gene is represented by a dot at its specific chromosomal coordinate, with

blue representing concordant copy number loss and gene downregulation and red representing concordant copy number amplification (copy number  $\geq 8$ ) and gene upregulation. Significance of concordant copy number/expression change is measured as a value of  $-\log_{10}(q\text{-value})$  times the sign of the direction of change. Dotted lines represent a significance threshold of  $-\log_{10}(q\text{-value}) = 0.05$  times the sign of the direction of change. Genes showing concordant copy number/expression changes and overlapping GISTIC peaks are listed above the plot. Asterisk denotes known PC oncogenes showing amplification but non-significant concordant copy number/expression change.

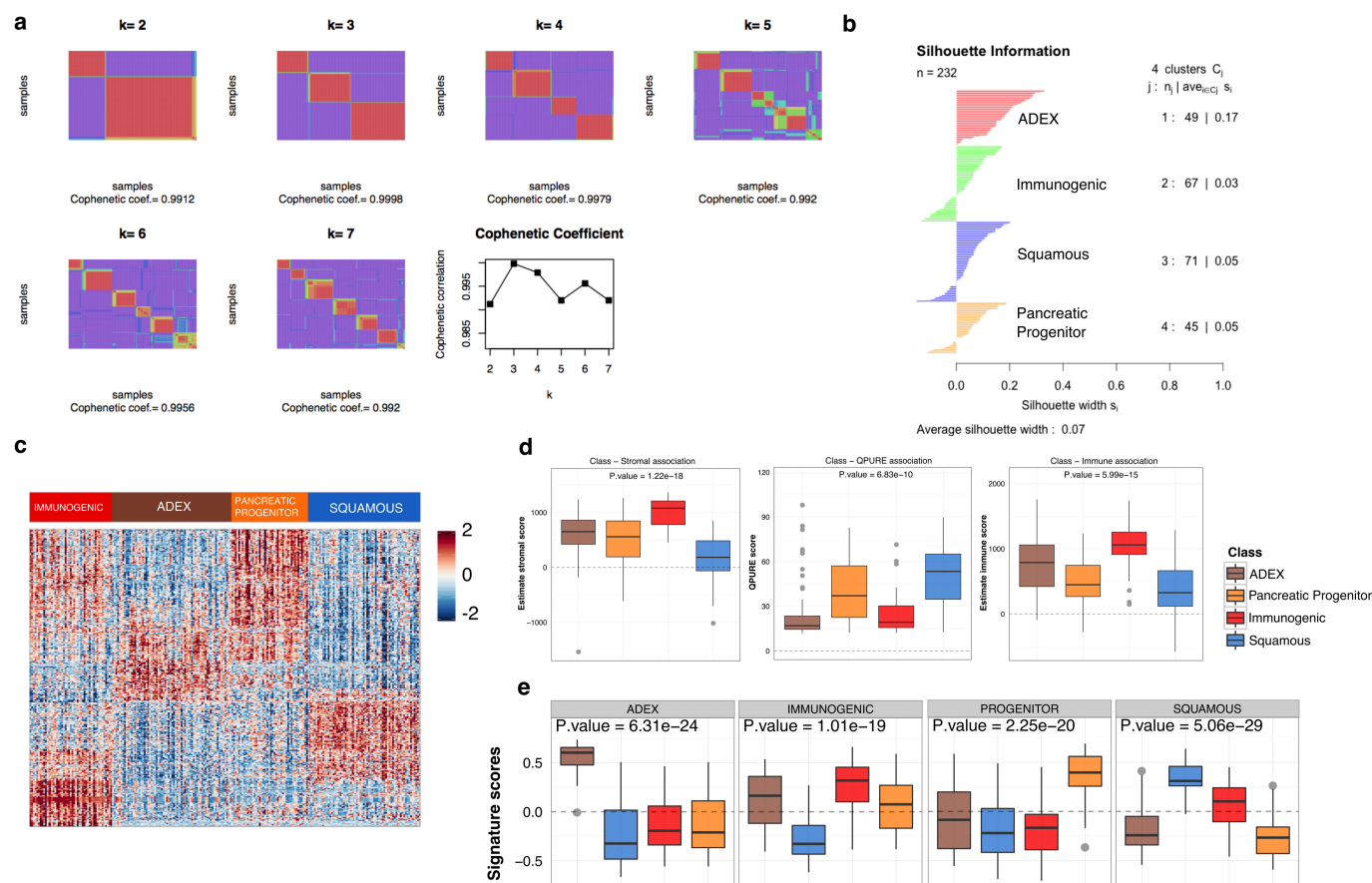




Extended Data Figure 3 | Classification of PC into 4 classes.

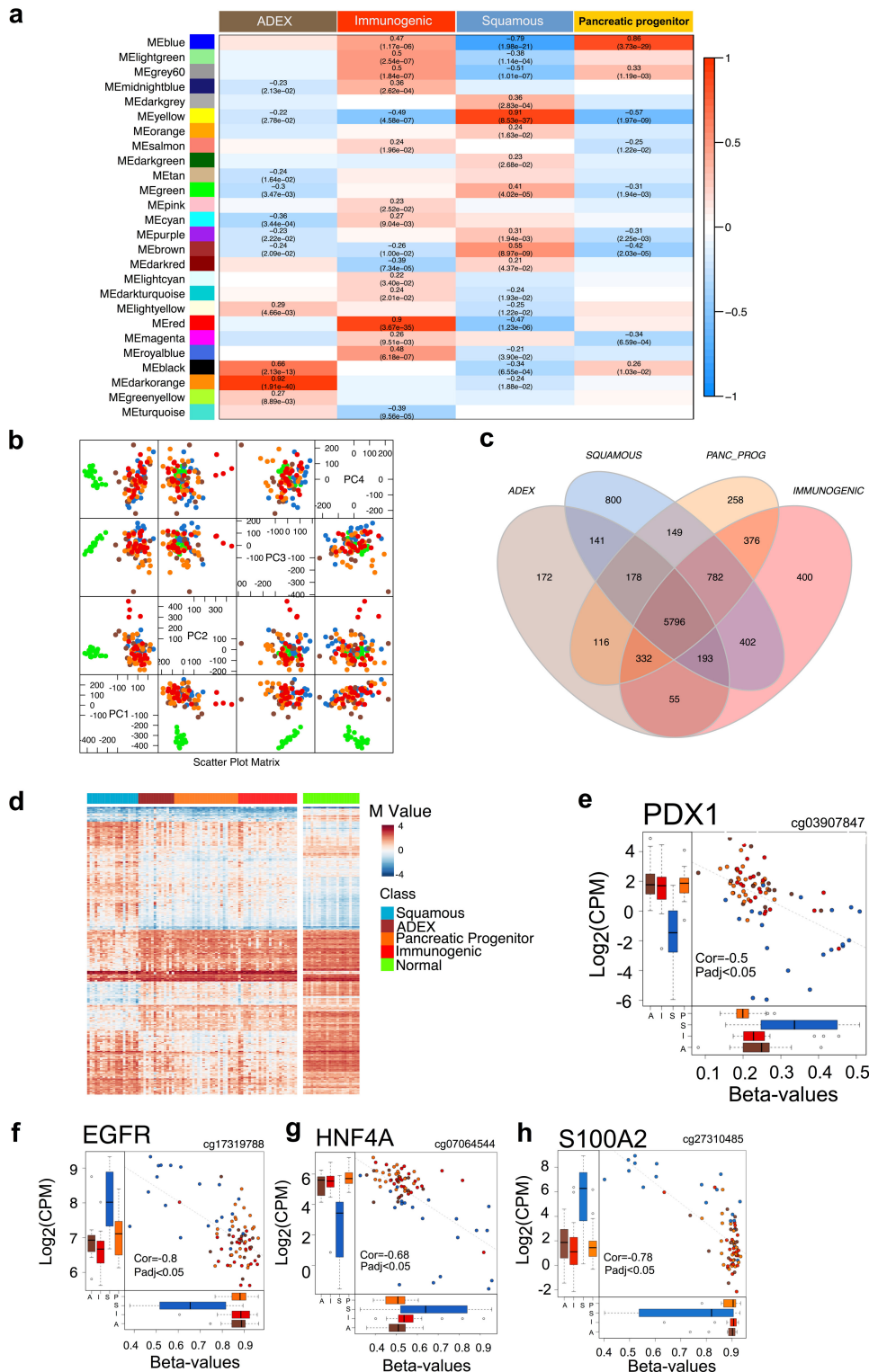
**a**, Unsupervised classification of PC RNAseq using NMF. Solutions are shown for  $k=2$  to  $k=7$  classes. A peak cophenetic correlation is observed for  $k=4$  classes. **b**, Silhouette information for  $k=4$  classes. **c–e**, Boxplots representing QPURE, stromal signature scores and immune signature scores stratified by class. Boxplots are annotated by a Kruskal–Wallis

$P$  value. For comparisons the following sample sizes were used: ADEX ( $n=16$ ); immunogenic ( $n=25$ ); squamous ( $n=25$ ); and pancreatic progenitor ( $n=30$ ). **f**, Heatmap showing differential gene expression between classes. Samples with positive silhouette widths were retained for ‘sam’ analysis. **g**, Heatmap showing overlap of the 4 classes identified in the current study and Collison *et al.* classification<sup>27</sup>.



**Extended Data Figure 4 | Identification of 4 robust PC classes in 232 PCs with mixed low and high cellularity.** **a**, Unsupervised classification of PC expression array data representing 232 samples using NMF. Solutions are shown for  $k = 2$  to  $k = 7$  classes. **b**, Silhouette information for  $k = 4$  classes. **c**, Heatmap showing differential gene expression between classes. **d**, Boxplots representing QPURE, stromal signature scores and immune signature scores stratified by class.

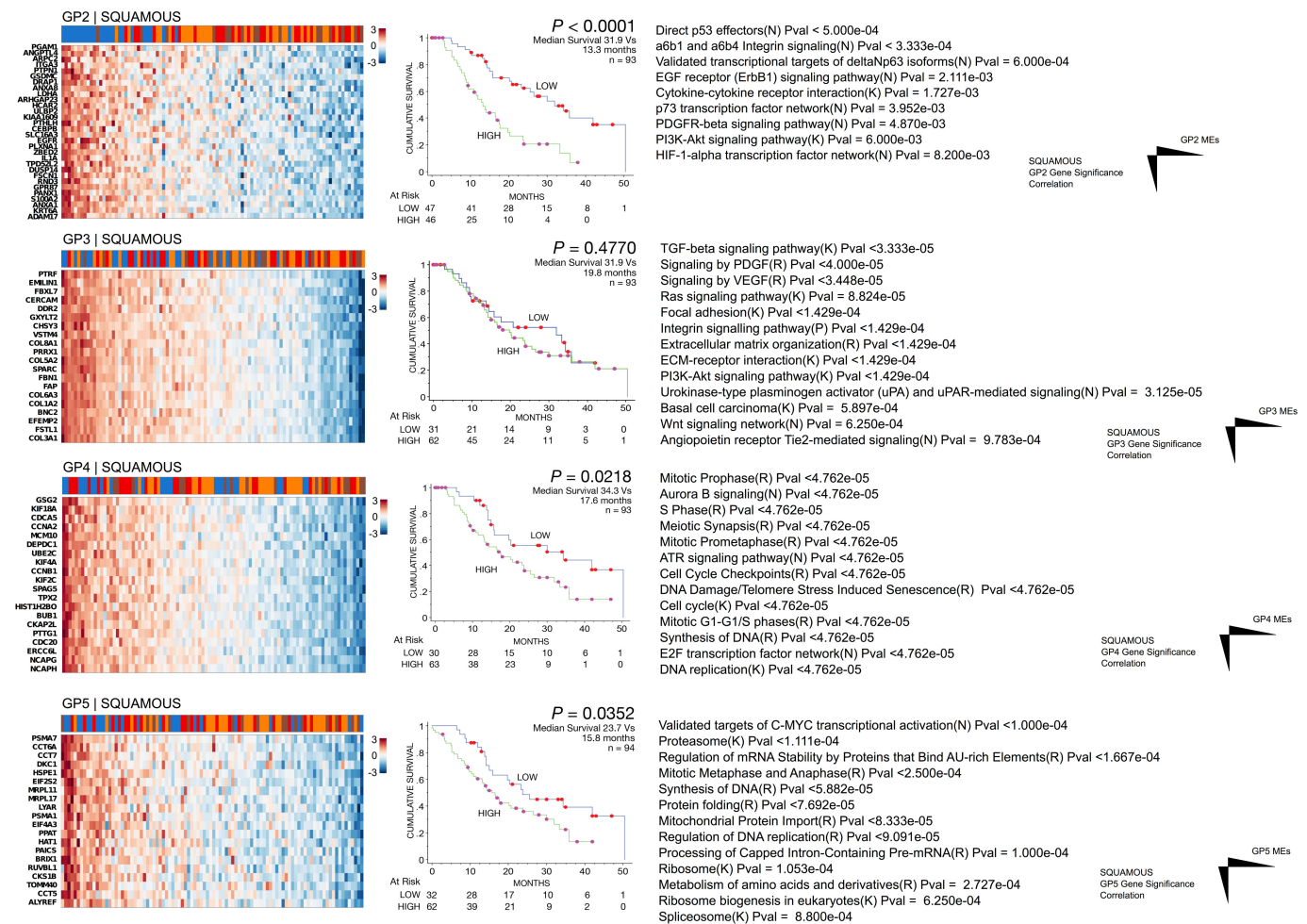
**e**, Boxplots representing ADEX, pancreatic progenitor, squamous and immunogenic signature scores defined using the RNA-seq PC set stratified by class. Boxplots in **d** and **e** are annotated by a Kruskal–Wallis  $P$  value. For comparisons the following sample sizes were used: ADEX ( $n = 49$ ); immunogenic ( $n = 67$ ); squamous ( $n = 71$ ); and pancreatic progenitor ( $n = 45$ ).



**Extended Data Figure 5 | Characterization of PC subtypes. a,** Heatmap showing the statistical significance of correlations observed between the expressions of genes significantly expressed in each PC class and gene programmes identified by WGCNA. Pearson correlations and Student's asymptotic  $P$  values are provided in each cell. **b,** Principal component analysis (PCA) using methylation data. Plot showing pairwise comparisons of samples distributed along the identified principle components (PC). Adjacent non-tumorous pancreatic samples represented as green points cluster as a distinct group. PC samples represented by points coloured brown (ADEX), blue (squamous), orange (pancreatic progenitor) and red (immunogenic) cluster together. **c,** Venn diagram showing the number of common and unique genes differentially methylated in the indicated PC subtypes when compared to adjacent non-tumorous pancreas. It

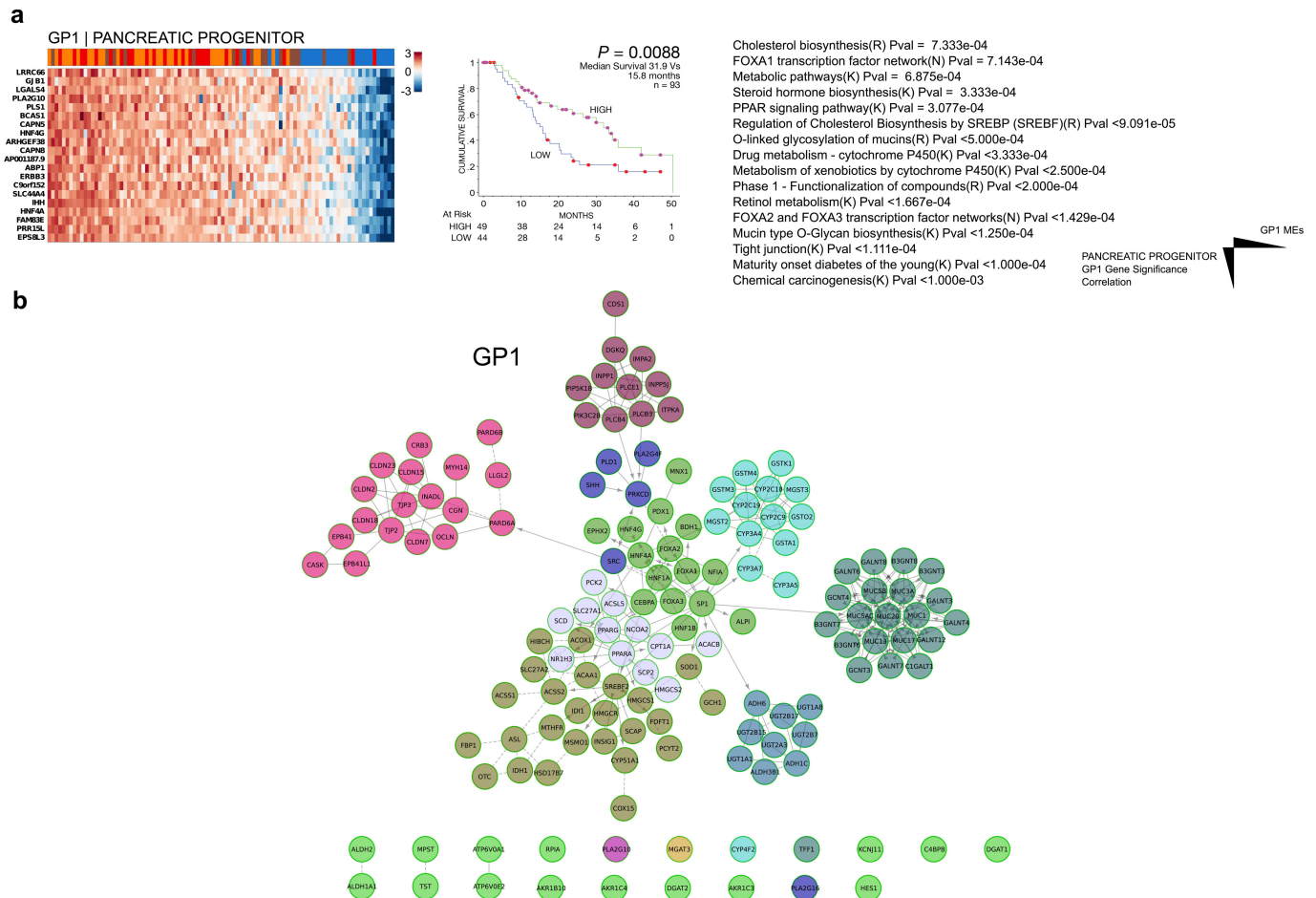
is observed that distinct subsets of genes are differentially methylated in the 4 PC subtypes. **d,** Heatmap showing genes that are significantly methylated between tumours comprising the squamous class and all other classes. Methylation values for the same genes in adjacent non-tumorous pancreas are also shown. **e–h,** Plots showing regulation of gene expression by methylation. Hyper- or hypomethylation of the indicated probe is associated with either the concordant downregulation or upregulation of the indicated gene. Pearson correlation and adjusted  $P$  values are provided for each gene methylation comparison. Boxplot colours designate class: ADEX (brown); immunogenic (red); squamous (blue); and pancreatic progenitor (orange). Single letter designations representing the first letter of each class are provided under the relevant boxes in each plot.





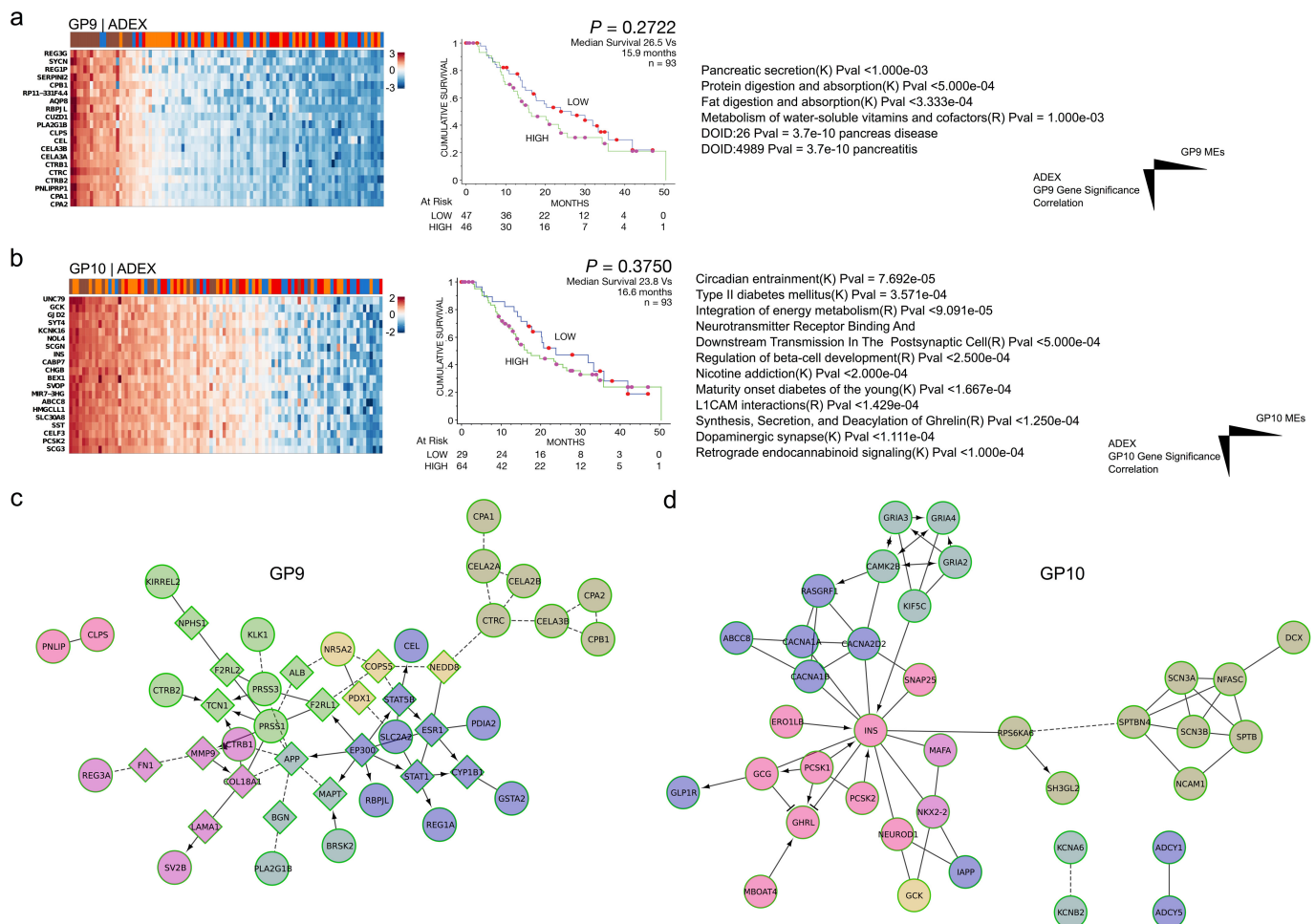
**Extended Data Figure 6 | Core gene programmes (GP) defining the squamous class.** Each panel shows from left to right: (i) a heatmap representing the genes in the specified gene programme most correlated with the indicated PC class with tumours ranked according to their gene programme module eigengene values (MEs) (PC classes are designated by colour as follows: ADEX (brown); pancreatic progenitor (orange);

immunogenic (red); and squamous (blue)); (ii) Kaplan–Meier analysis comparing survival of patients having either high or low gene programme MEs; (iii) pathways significantly enriched in a given GP functional plugin. *P* values represent FDR < 0.05.



**Extended Data Figure 7 | Gene programme defining the pancreatic progenitor class.** **a**, Panel showing from left to right: (i) a heatmap representing the genes in GP1 most correlated with the pancreatic progenitor class with tumours ranked according to their GP1 module eigengene values (MEs); (ii) Kaplan–Meier analysis comparing survival of

patients having either high or low GP1 MEs; (iii) pathways significantly enriched in a GP1 FI sub-network defined by the ReactomeFI Cytoscape plugin.  $P$  values represent FDR <0.05. **b**, Network diagram depicting pathways significantly enriched in GP1 (FDR <0.0001). Different node colours indicate different network clusters or closely interconnected genes.



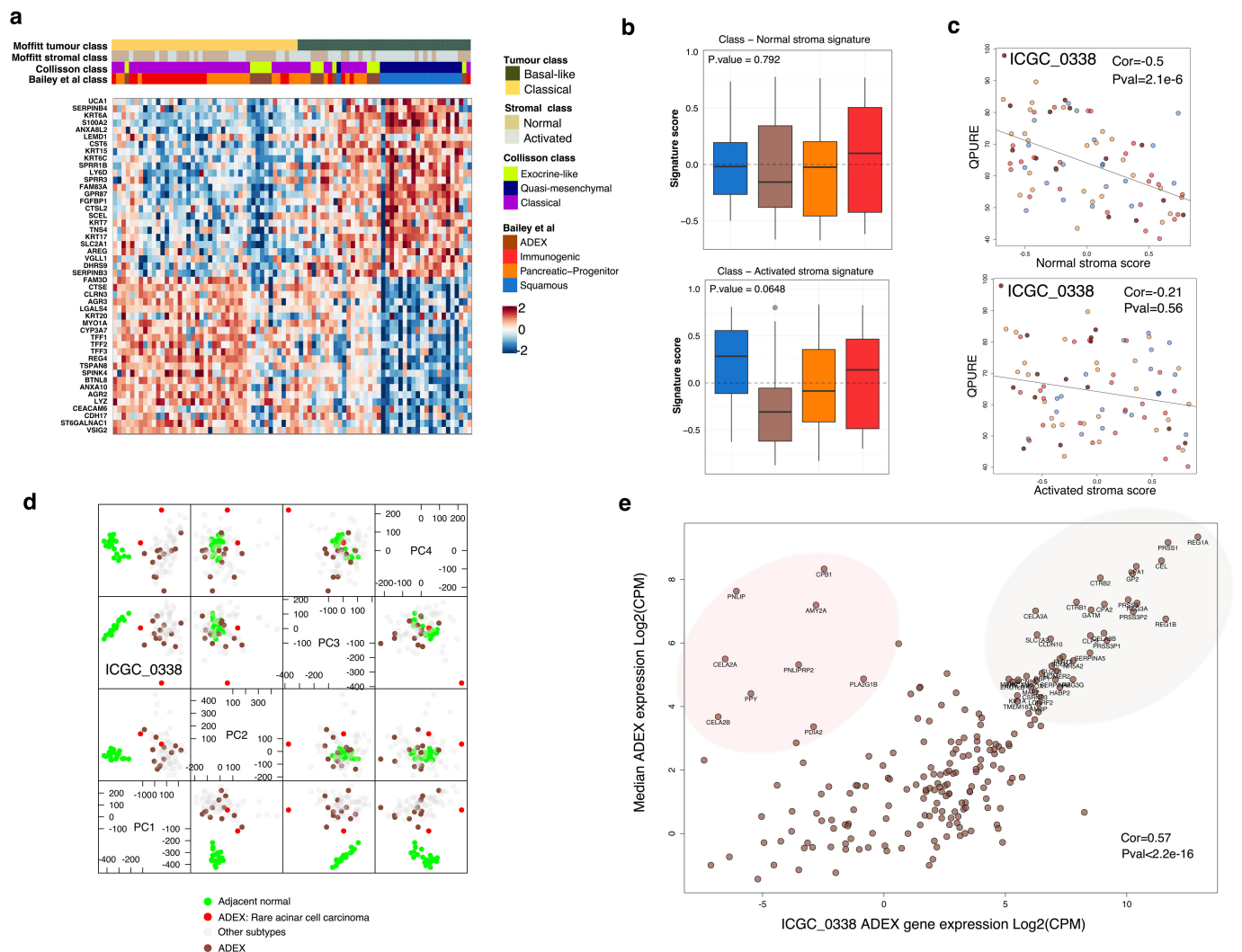
### Extended Data Figure 8 | Gene programmes defining the ADEX class.

**a, b**, Panel showing from left to right: (i) a heatmap representing the genes in the specified GP most correlated with the ADEX class with tumours ranked according to their GP module eigengene values (MEs); (ii) Kaplan-Meier analysis comparing survival of patients having either high or low GP MEs; (iii) pathways significantly enriched in a GP FI sub-network defined by the ReactomeFI Cytoscape plugin.  $P$  values represent FDR <0.05.

**c**, Network diagram depicting pathways significantly enriched in GP9

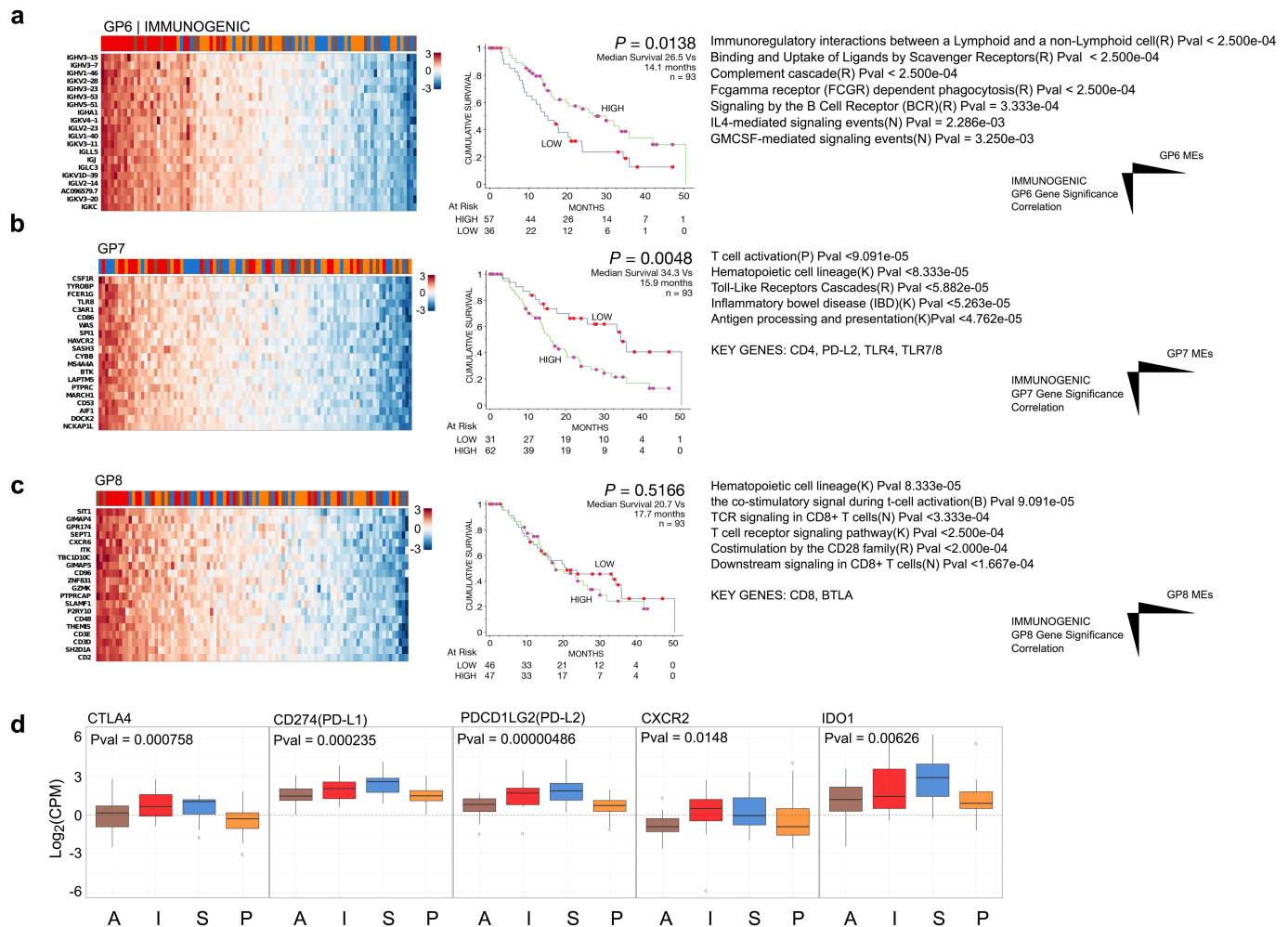
(FDR <0.0001). Different node colours indicate different network clusters or closely interconnected genes. Genes comprising GP9 are indicated as coloured circles, whereas linker genes (genes not comprising GP9 but forming multiple connections in the network) are indicated as coloured diamonds. **d**, Network diagram depicting pathways significantly enriched in GP10 (FDR <0.0001). Different node colours indicate different network clusters or closely interconnected genes.





**Extended Data Figure 9 | Stratification of PC RNASeq data according to Moffitt *et al.*** **a**, Heatmap showing the stratification of the PC cohort of the current study using the tumour subtype classifier published in Moffitt *et al.*<sup>28</sup>. PCs were classified by consensus clustering using the top 50 weighted genes associated with the basal-like or classical subtypes. **b**, Boxplots showing the distribution of normal and activated stroma signature scores between the 4 PC classes identified in the current study. Boxplots are annotated by a Kruskal–Wallis *P* value. A significant difference in activated stroma signature scores was observed between squamous and ADEX tumours *P* value < 0.01 (*t*-test). Boxplot colours designate class: ADEX (brown); immunogenic (red); squamous (blue); and pancreatic progenitor (orange). **c**, Plots showing correlation between tumour cellularity, presented as a QPURE score, and either activated or normal stroma signature scores. Plots are annotated with Pearson correlation scores and significance values, with a linear fit represented by a solid line. Sample ICGC\_0338, a rare acinar cell carcinoma is highlighted.

This sample exhibits near 100% cellularity and has low activated or normal stroma signature scores. **d**, Principal component analysis (PCA) using methylation data. Plot showing pairwise comparisons of samples distributed along the identified principle components (PC). Adjacent non-tumorous pancreatic samples represented as green points cluster as a distinct group relative to ADEX samples (brown and red points). Rare acinar cell carcinomas (red) cluster with other ADEX samples (brown). All other PC samples are shown as grey points. **e**, Plot showing the correlation of expression of representative genes expressed in acinar cell carcinoma sample ICGC\_0338 compared to the median expression of the same genes across all other ADEX samples. A red shaded region encompasses genes showing high median expression in all other ADEX but low expression in ICGC\_0338. A brown shaded region encompasses genes showing high median expression in all other ADEX and correlatively high expression in ICGC\_0338. Pearson's correlation and significance are indicated.



**Extended Data Figure 10 | Gene programmes defining the immunogenic class.** **a–c**, Each panel shows from left to right: (i) a heatmap representing the genes in the specified gene programme most correlated with the indicated PC class with tumours ranked according to their gene programme module eigengene values (MEs). PC classes are designated by colour as follows: ADEX (brown); pancreatic progenitor (orange); immunogenic (red); and squamous (blue); (ii) Kaplan–Meier analysis comparing survival of patients having either high or low gene programme MEs; (iii) pathways significantly enriched

in a given GP functional interaction (FI) sub-network defined by the ReactomeFI Cytoscape plugin. Corresponding Cytoscape files comprising GP ReactomeFI subnetworks are provided. **d**, Boxplot of immune gene expression stratified by class. Boxplots are annotated by a Kruskal–Wallis *P* value and box colours designate class: ADEX (brown); immunogenic (red); squamous (blue); and pancreatic progenitor (orange). Single letter designations representing the first letter of each class are provided under the relevant boxes in each plot.

# High-fat diet enhances stemness and tumorigenicity of intestinal progenitors

Semir Beyaz<sup>1,2\*</sup>, Miyeko D. Mana<sup>1\*</sup>, Jatin Roper<sup>1,3\*</sup>, Dmitriy Kedrin<sup>1,4</sup>, Assieh Saadatpour<sup>5</sup>, Sue-Jean Hong<sup>6</sup>, Khristian E. Bauer-Rowe<sup>1</sup>, Michael E. Xifaras<sup>1</sup>, Adam Akkad<sup>1</sup>, Erika Arias<sup>1</sup>, Luca Pinello<sup>5</sup>, Yarden Katz<sup>7</sup>, Shweta Shinagare<sup>1</sup>, Monther Abu-Remaih<sup>1,6</sup>, Maria M. Mihaylova<sup>1,6</sup>, Dudley W. Lamming<sup>8</sup>, Rizkullah Dogum<sup>1</sup>, Guoji Guo<sup>2</sup>, George W. Bell<sup>6</sup>, Martin Selig<sup>4</sup>, G. Petur Nielsen<sup>4</sup>, Nitin Gupta<sup>9</sup>, Cristina R. Ferrone<sup>4</sup>, Vikram Deshpande<sup>4</sup>, Guo-Cheng Yuan<sup>5</sup>, Stuart H. Orkin<sup>2</sup>, David M. Sabatini<sup>1,6,7</sup> & Ömer H. Yilmaz<sup>1,4,7</sup>

**Little is known about how pro-obesity diets regulate tissue stem and progenitor cell function. Here we show that high-fat diet (HFD)-induced obesity augments the numbers and function of *Lgr5*<sup>+</sup> intestinal stem cells of the mammalian intestine. Mechanistically, a HFD induces a robust peroxisome proliferator-activated receptor delta (PPAR- $\delta$ ) signature in intestinal stem cells and progenitor cells (non-intestinal stem cells), and pharmacological activation of PPAR- $\delta$  recapitulates the effects of a HFD on these cells. Like a HFD, *ex vivo* treatment of intestinal organoid cultures with fatty acid constituents of the HFD enhances the self-renewal potential of these organoid bodies in a PPAR- $\delta$ -dependent manner. Notably, HFD- and agonist-activated PPAR- $\delta$  signalling endow organoid-initiating capacity to progenitors, and enforced PPAR- $\delta$  signalling permits these progenitors to form *in vivo* tumours after loss of the tumour suppressor *Apc*. These findings highlight how diet-modulated PPAR- $\delta$  activation alters not only the function of intestinal stem and progenitor cells, but also their capacity to initiate tumours.**

The mammalian intestine is known to respond to dietary signals<sup>1</sup>. *Lgr5*<sup>+</sup> intestinal stem cells (ISCs) remodel intestinal composition in response to diet-induced cues by adjusting their production of daughter stem cells and (non-ISC, transit-amplifying cells) progenitor cells, the latter of which differentiate into the diverse cell types of the intestine<sup>1–3</sup>. The *Lgr5*<sup>+</sup> ISCs reside at the base of intestinal crypts adjacent to Paneth cells, which are a central component of the ISC niche and regulate stem-cell biology in response to calorie-restricted diets<sup>1,4</sup>.

Although important epidemiological and rodent studies link obesity to colon cancer incidence<sup>3,5–7</sup>, little is known about how the adaptation of stem and progenitor cells to pro-obesity diets alters the potential of these cells to initiate tumours<sup>3</sup>. In the mouse intestine, *Lgr5*<sup>+</sup> ISCs serve as the cell-of-origin for the precancerous adenomatous lesions caused by loss of the *Apc* tumour suppressor gene; yet, it is unclear whether this occurs in the context of obesity-linked intestinal tumorigenesis<sup>8,9</sup>. Here, we interrogate how long-term HFD-induced obesity influences intestinal stem and progenitor cell function and the cellular origins of intestinal dysplasia.

## HFD boosts ISC counts and crypt function

To assess the effects of obesity on intestinal homeostasis, we maintained mice on a long-term HFD (60% fat diet; Extended Data Fig. 1o) for 9–14 months, which is sufficient to observe many of the metabolic phenotypes associated with obesity<sup>10,11</sup>. Consistent with previous reports, HFD-fed mice gained considerably more mass than their standard chow-fed counterparts (Extended Data Fig. 1a). While the small intestines from HFD-fed mice were shorter in length (Extended Data Fig. 1c) and weighed less (Extended Data Fig. 1b), there was no change in the density of crypt-villous units (Extended Data Fig. 1d) or the

number of apoptotic cells (Extended Data Fig. 1n). Morphologically, a HFD led to a mild reduction in villi length (Extended Data Fig. 1g), an associated decrease in villous enterocyte numbers (Extended Data Fig. 1f), and an increase in crypt depth (Extended Data Fig. 1e). A HFD did not change the numbers of chromogranin A<sup>+</sup> enteroendocrine cells or Alcian blue<sup>+</sup> goblet cells per crypt-villus unit of the small intestine (Extended Data Fig. 2a–d).

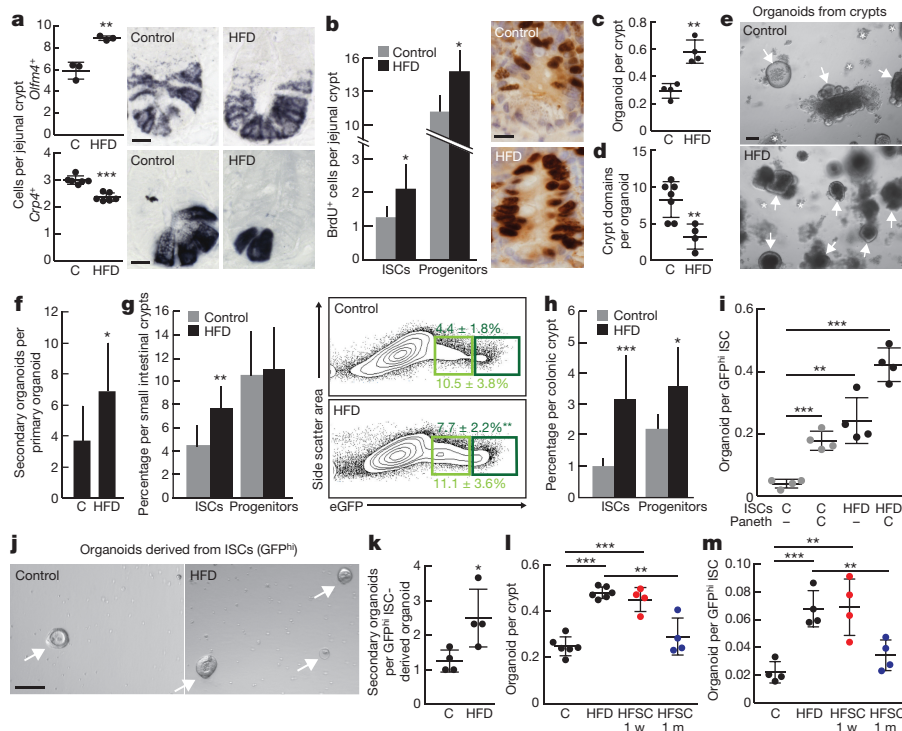
To address how a HFD affects the frequency of ISCs, we performed *in situ* hybridization for olfactomedin 4 (*Olfm4*), a marker expressed by the *Lgr5*<sup>+</sup> ISCs<sup>12</sup>. Compared to mice fed a standard chow diet, those on a HFD had a 50% increase in the number of *Olfm4*<sup>+</sup> ISCs (Fig. 1a and Extended Data Fig. 1l). By contrast, a HFD reduced cryptdin 4<sup>+</sup> (*Crp4*<sup>+</sup>, also known as *Defa4*<sup>+</sup>) niche Paneth cell numbers by 23% (Fig. 1a and Extended Data Fig. 1m). These observations lead to two conclusions: first, a HFD enhances ISC numbers and self-renewal (for example, deeper crypts with more *Olfm4*<sup>+</sup> ISCs) at the expense of differentiation (shorter and less cellular villi); and, second, the increase in ISCs occurs despite a reduction in Paneth cell numbers, raising the possibility that under a HFD, ISCs adjust to fewer interactions from their Paneth cell niche.

Given that ISC numbers and proliferation (Fig. 1b, Extended Data Figs 1h–k and Supplementary Information) increase in a HFD, we asked whether a HFD also boosts intestinal regeneration. Using an *in vitro* approach, we assessed the ability of isolated intestinal crypts to form organoid bodies in 3-D culture. These organoids recapitulate the epithelial architecture and cellular diversity of the mammalian intestine and are a proxy for ISC activity, as only stem cells can initiate and maintain these structures long-term<sup>1,13</sup>. HFD-derived crypts from the small intestine and colon were more likely to initiate mini-intestines in

<sup>1</sup>The David H. Koch Institute for Integrative Cancer Research at MIT, Department of Biology, MIT, Cambridge, Massachusetts 02139, USA. <sup>2</sup>Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Howard Hughes Medical Institute, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>Division of Gastroenterology and Molecular Oncology Research Institute, Tufts Medical Center, Boston, Massachusetts 02111, USA. <sup>4</sup>Departments of Pathology, Gastroenterology, and Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>5</sup>Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, USA. <sup>6</sup>Whitehead Institute for Biomedical Research, Howard Hughes Medical Institute, Department of Biology, MIT, Cambridge, Massachusetts 02142, USA. <sup>7</sup>Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. <sup>8</sup>Department of Medicine, University of Wisconsin-Madison, Madison, Wisconsin 53705, USA. <sup>9</sup>Division of Digestive Diseases, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA.

\*These authors contributed equally to this work.





**Figure 1 | HFD augments ISC numbers and function.** **a**, Quantification of *Olfm4*<sup>+</sup> ISCs ( $n = 3$ ) and *Crp4*<sup>+</sup> Paneth cells ( $n = 6$ ) in the proximal jejunum of control (C) and HFD-fed mice by *in situ* hybridization. **b**, BrdU incorporation in ISCs (crypt base columnar cells) and progenitors (transit-amplifying cells) after a 4-h pulse ( $n = 6$ ). **c–e**, Organoid per crypt (**c**,  $n = 4$ ) and crypt domain (**d**,  $n = 7$ ) quantification from control and HFD-fed mice (**d**,  $n = 4$ ). Representative images: day-7 organoids (**e**). Arrows denote organoids; asterisks denote aborted crypts. **f**, Number of secondary organoids per dissociated crypt-derived primary organoids ( $n = 9$  primary organoids, 3 primary organoids per sample were individually subcloned in 3 independent experiments). **g, h**, Frequencies of ISCs (*Lgr5*-GFP<sup>hi</sup>, dark green) and progenitors (*Lgr5*-GFP<sup>low</sup>, light green) in the entire small intestine (**g**,  $n = 10$ ) and colon (**h**,  $n = 8$ ) as measured by flow cytometry.

culture than those from controls (Fig. 1c, e and Extended Data Fig. 3j). Furthermore, these organoids were more cystic (that is, less differentiated<sup>14</sup>) in structure and contained fewer crypt domains (Fig. 1d). When sub-cloned, HFD-derived primary organoids generated more secondary organoids (Fig. 1f and Extended Data Fig. 3k). Consistent with these findings, HFD crypt-derived organoids had higher frequencies of *Lgr5*<sup>+</sup> ISCs compared to controls (Extended Data Fig. 4a, d, e), and possessed diverse intestinal cell types, such as Paneth cells, ISCs, enteroendocrine cells and goblet cells (Extended Data Fig. 4b–f).

To determine whether a HFD also augments crypt regeneration *in vivo*, we performed a clonogenic microcolony assay to test for ISC activity<sup>1,15</sup>. After administration of a lethal dose of irradiation, HFD-fed mice manifested increased numbers of surviving, proliferating crypts (*Ki67*<sup>+</sup> cells per crypt) that possessed more *Olfm4*<sup>+</sup> ISCs per unit length of intestine relative to controls (Extended Data Fig. 2e–g). These data support the notion that a HFD boosts the numbers and regenerative capacity of ISCs *in vitro* and *in vivo*.

### HFD reduces the niche dependence of ISCs

To assess the effects of a HFD on ISCs and progenitors, we used *Lgr5-EGFP-IRES-CreERT2* knock-in mice for the quantification and isolation of green fluorescent protein (GFP)-expressing ISCs (*Lgr5*-GFP<sup>hi</sup>) and progenitor cells (*Lgr5*-GFP<sup>low</sup>)<sup>2</sup>. Compared to controls, mice on a HFD had an increased frequency of *Lgr5*-GFP<sup>hi</sup> ISCs in the small intestine (Fig. 1g) and colon (Fig. 1h and Extended Data Fig. 3g).

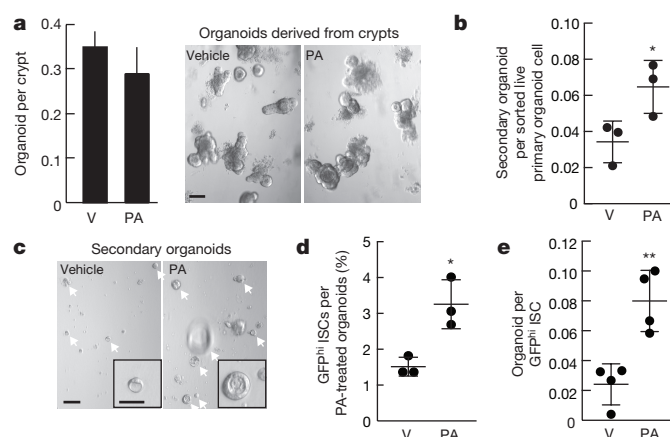
The opposing effects of a HFD on ISC and Paneth cell numbers led us to ask whether a HFD alters ISC function and niche dependence.

**i, j**, Organoid-initiating capacity of control and HFD ISCs cultured with/without Paneth cells (**i**,  $n = 4$ ). Representative images: day-5 primary organoids (arrows, **j**). **k**, Number of secondary organoids per dissociated ISC-derived primary organoid ( $n = 4$ ). **l, m**, Crypts (**l**) and ISCs (**m**) isolated from HFD-fed mice that were reverted to a standard chow diet (HFSC) retained augmented organoid-forming capacity for 1 week (w) (red;  $n = 4$ ) but not for 1 month (m) (blue;  $n = 4$ ) when compared to their HFD counterparts ( $n = 6$  crypts,  $n = 4$  ISCs). Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (Student's *t*-tests). Scale bars, 20  $\mu$ m (**a, b**) and 100  $\mu$ m (**e, j**). Histological analysis: **a**, *Olfm4*: 10 crypts per group, *Crp4*: 50 crypts per group; **b**, 50 crypts per group in each experiment.

We assayed the clonogenic potential of ISCs from control and HFD-fed mice either alone or in combination with the niche Paneth cells<sup>1</sup>. Consistent with earlier studies<sup>1,4,13</sup>, control ISCs by themselves inefficiently formed organoids, but robustly formed organoids when co-cultured with Paneth cells (Fig. 1i). Surprisingly, HFD-derived ISCs alone (without Paneth cells) had an increased capacity to initiate organoids with multilineage differentiation and more secondary organoids than control ISCs (Fig. 1i–k and Extended Data Fig. 4h, i, l, m). Co-culture with Paneth cells further increased the organoid-initiating activity of HFD ISCs (Fig. 1i). Organoids derived from control and HFD ISCs alone effectively produced Paneth cells within 24 h of culture (Extended Data Fig. 4j, k). Furthermore, crypts and ISCs isolated from mice that had been on a HFD, but were returned to a standard chow diet, retained an enhanced capacity to initiate organoids for more than 7 days but less than 4 weeks, indicating that the effects of a HFD are reversible (Fig. 1l, m). These data, together with the observation that a HFD uncouples the *in vivo* expansion of ISCs from their Paneth cell niche, suggest ISCs undergo autonomous changes in response to a HFD that poise them for niche-independent growth in the organoid assay.

### Fatty acids drive organoid self-renewal

To address whether dietary constituents of the HFD can recapitulate aspects of the HFD-evoked stem-cell phenotype, we expanded control organoids in crypt media supplemented with palmitic acid, a main component of the HFD<sup>16</sup>. Treatment with palmitic acid did not alter the clonogenic potential of control crypts in primary culture



**Figure 2 | *Ex vivo* exposure of intestinal organoids to palmitic acid recapitulates aspects of a HFD.** **a**, Clonogenicity of naive crypts cultured with 30  $\mu$ M palmitic acid (PA) in primary organoid cultures ( $n = 5$ ). Representative images: day-4 organoids. V, vehicle. **b**, **c**, Secondary organoid formation of 1,000 sorted live primary organoid cells after 4 weeks of 30  $\mu$ M palmitic acid treatment (**b**,  $n = 3$ ). Representative images: day-4 secondary organoids (arrows, **c**). **d**, **e**, Frequency (**d**) and organoid initiation (**e**) of ISCs ( $Lgr5$ -GFP<sup>hi</sup>) after 4 weeks of 30  $\mu$ M palmitic acid exposure (**d**,  $n = 3$ ; **e**,  $n = 4$ ). Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$  (Student's  $t$ -tests). Scale bars, 100  $\mu$ m (**a**, **c**) and 50  $\mu$ m (**c**, inset).

(Fig. 2a). However, as observed with organoids from HFD-fed mice, primary organoids exposed *ex vivo* to palmitic acid gave rise to more secondary organoids than controls (Fig. 2b, c and Extended Data Fig. 5a). Consistent with these findings, organoids treated with palmitic acid possessed nearly twofold more  $Lgr5$ <sup>+</sup> ISCs (Fig. 2d and Extended Data Fig. 5b), and manifested reduced niche dependence in the organoid assay (Fig. 2e). Similar results were obtained with other fatty acids such as oleic acid and a lipid mixture in mouse and human intestinal organoids (Fig. 3h–k and Extended Data Fig. 5c–f). These findings indicate that key dietary constituents of a HFD are sufficient to recapitulate aspects of the *in vivo* HFD stem-cell phenotype.

### HFD acts through PPAR- $\delta$ in ISCs

To gain mechanistic insight into how HFD mediates these effects, we performed messenger RNA sequencing on isolated  $Lgr5$ -GFP<sup>hi</sup> ISCs and  $Lgr5$ -GFP<sup>low</sup> progenitor cells from control and HFD-fed mice, respectively (Extended Data Fig. 6p). Gene set enrichment analysis (GSEA) pathway and transcription factor binding motif analyses revealed enrichment for transcriptional targets and binding motifs of the nuclear receptor peroxisome proliferator-activated receptor (PPAR) family and PPAR heterodimeric binding partners liver/retinoid X receptor<sup>17–19</sup> (LXR/RXR; Extended Data Fig. 6c, d). Three members ( $\alpha$ ,  $\delta$  and  $\gamma$ ) comprise the PPAR family<sup>17</sup>; among these, PPAR- $\delta$  (*Ppard*) is the predominant one expressed in intestinal stem and progenitor cells at the mRNA level in control and HFD-fed mice (Extended Data Fig. 6a, b). Therefore, we focused our attention on PPAR- $\delta$  and its potential role in coupling a HFD to ISC function.

Although PPAR- $\delta$  expression itself did not substantially increase (Extended Data Fig. 6a, b), the HFD robustly induced expression of many of its target genes at the mRNA levels in both the small intestine (Extended Data Fig. 6e) and colon (Extended Data Fig. 3h). The induction of the PPAR- $\delta$  program was verified at the protein level in ISCs and progenitors (Fig. 3a). To address functionally whether engagement of a PPAR- $\delta$  program mimics the HFD, we administered the PPAR- $\delta$  agonist GW501516 for 4 weeks to *Lgr5*-EGFP-IRES-CreERT2 mice<sup>20,21</sup>. Treatment led to strong induction of PPAR- $\delta$  target proteins in ISCs and progenitors (Fig. 3a). Furthermore, agonist-activated PPAR- $\delta$  signalling augmented the *in vivo* frequencies of *Olfr4*<sup>+</sup> and  $Lgr5$ <sup>+</sup> ISCs (Fig. 3b, e and Extended Data Fig. 6f) and proliferation of stem and

progenitors cells (Fig. 3c), but had no effect on Paneth cell numbers (Fig. 3b and Extended Data Fig. 6g). Notably, small intestinal (Fig. 3d) and colonic (Extended Data Fig. 3l, m) crypts from agonist-treated mice initiated more organoids than those from vehicle-treated mice. Similar to ISCs from HFD-fed mice, ISCs derived from agonist-treated mice were more effective at Paneth cell-independent organoid-initiation than their control counterparts (Fig. 3f). In addition, organoids exposed to the PPAR- $\delta$  agonist had more  $Lgr5$ <sup>+</sup> ISCs (Fig. 3g) and more self-renewing capacity in secondary assays (Fig. 3i). These data indicate that sustained PPAR- $\delta$  signalling largely recapitulates the effects of a HFD on ISC function.

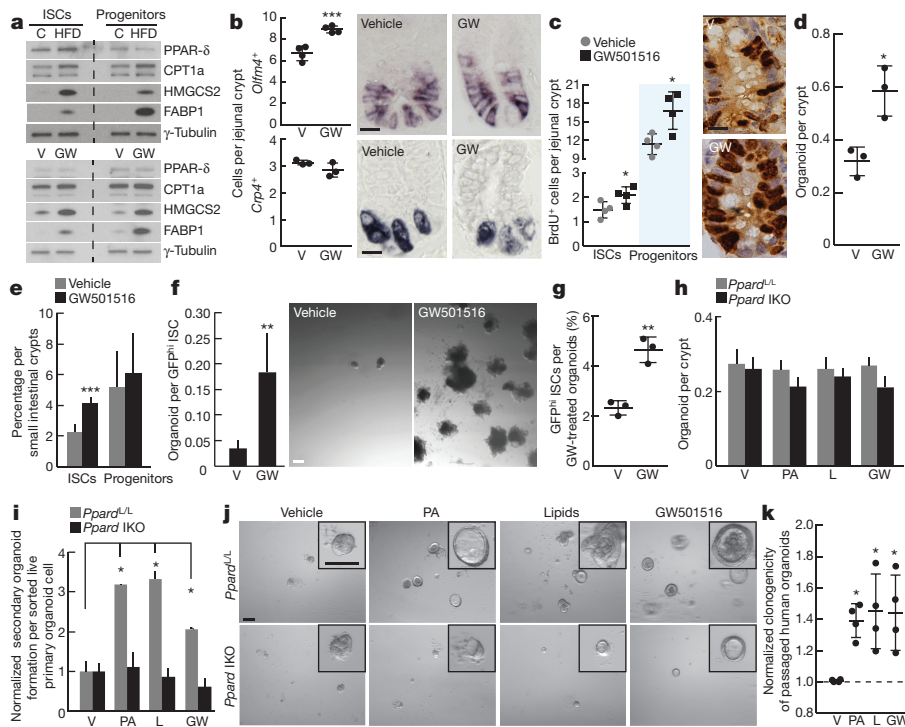
Because *ex vivo* fatty acids mimic aspects of a HFD, we asked whether this phenomenon occurs through PPAR- $\delta$  signalling. Like a HFD, we observed that *ex vivo* exposure of mouse and human organoids to fatty acids evokes a robust PPAR- $\delta$  program (Extended Data Figs 5g–j and 6h). To assess the necessity of PPAR- $\delta$  in this response to fatty acids, we generated tamoxifen-inducible, intestine-specific *Ppard* conditional mice (Extended Data Fig. 6i, j). Acute ablation of *Ppard* in the intestine had no noticeable effects on the numbers, proliferation or function of ISCs and progenitor cells (Fig. 3h and Extended Data Fig. 6i–n). However, loss of *Ppard* blocked both the self-renewal enhancing effects of fatty acids and PPAR- $\delta$  agonist (Fig. 3i, j), as well as the induction of PPAR- $\delta$  target gene expression in secondary organoid assays (Extended Data Fig. 6o). These findings demonstrate that PPAR- $\delta$  mediates fatty-acid-driven organoid self-renewal.

### HFD and PPAR- $\delta$ raise $\beta$ -catenin activity

Because HFD and PPAR- $\delta$  activation confer increased stem-cell function, we asked whether these interventions regulate the Wnt/ $\beta$ -catenin pathway, which is required for ISC maintenance<sup>2,22</sup>. First, we observed more nuclear  $\beta$ -catenin, a proxy for its activity, in sorted ISCs and progenitors and on intestinal sections from HFD and PPAR- $\delta$  agonist-treated mice compared to controls (Extended Data Fig. 7c–i). Second, crypts from HFD and PPAR- $\delta$  agonist-treated mice required less exogenous Wnt for organoid maintenance than controls (Fig. 4a, b). Lastly, we found that increased levels of nuclear  $\beta$ -catenin associate with PPAR- $\delta$  in HFD crypts (Extended Data Fig. 7j–l).

To address how a HFD and agonist-activated PPAR- $\delta$  influence  $\beta$ -catenin transcriptional activity, we performed microfluidic-based multiplexed single-cell quantitative reverse transcription PCR (qRT-PCR) using primers for a curated list of known  $\beta$ -catenin target genes that includes ISC markers (Supplementary Table 2 and Extended Data Fig. 8a, d). While a HFD did not alter expression of stem-cell signature genes (that is, *Lgr5*) that differ between stem and progenitor cells<sup>23,24</sup> (Extended Data Fig. 8b, e, g, h), it evoked expression of a subset of  $\beta$ -catenin target genes such as *Bmp4*, *Jag1*, *Jag2* and *Edn3* in ISCs and progenitors (Fig. 4c, e, Extended Data Figs 3i and 8c, f, i, j and Supplementary Information). Single-cell qRT-PCR analysis confirmed that agonist-activated PPAR- $\delta$  also induced transcription of *Bmp4*, *Jag1*, *Jag2* and *Edn3* in ISCs and progenitors (Fig. 4d, f). We further validated *Jag1* expression by single-molecule *in situ* hybridization and found that it was broadly expressed within HFD (Extended Data Fig. 8k) and PPAR- $\delta$  agonist-treated crypts (Extended Data Fig. 8l). Moreover, in response to *in vitro* fatty acids, PPAR- $\delta$  was required for the induction of *Jag1* and *Jag2* in secondary organoids (Extended Data Fig. 6o). Collectively, these results support a model in which a HFD activates a PPAR- $\delta$ -mediated subset of  $\beta$ -catenin target genes in ISCs and progenitor cells.

To interrogate whether a similar program exists in an alternative model of obesity, we assessed how the intestine adapts to obesity in leptin receptor deficient (*db/db*) mice—an obesity model that develops on a standard diet. Overall, we found that intestinal adaptation in the *db/db* obesity model was mostly opposite to what we observed in HFD-fed mice (Extended Data Fig. 9 and Supplementary Information). Such differences highlight that, even in obesity, diet affects ISC and progenitor biology.



**Figure 3 | Activated PPAR- $\delta$  in ISCs mediates the effects of a HFD.**

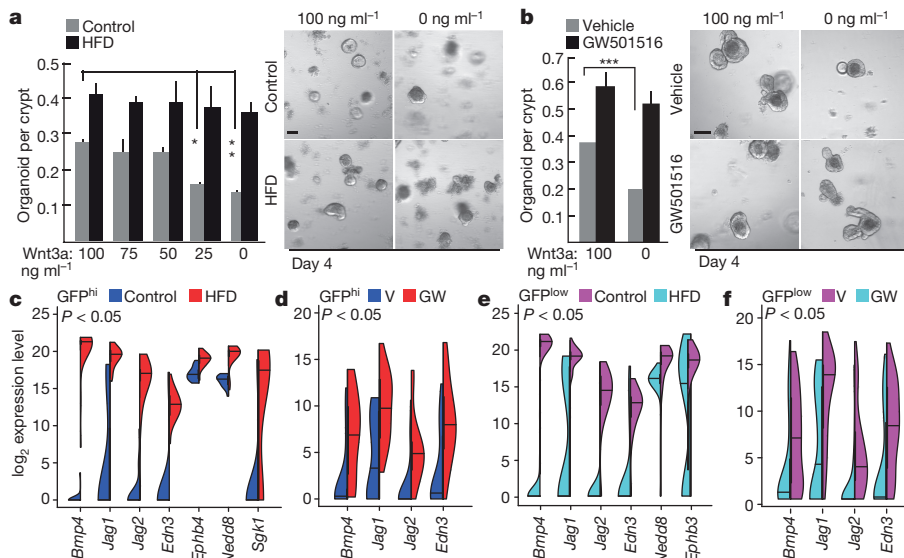
**a**, Immunoblots of PPAR- $\delta$  target proteins in flow-sorted ISCs (Lgr5-GFP<sup>hi</sup>) and progenitors (Lgr5-GFP<sup>low</sup>) from control, HFD, vehicle and GW501516 (GW) mice ( $n = 2$ ). **b**, Quantification of *Olfm4*<sup>+</sup> ISCs ( $n = 4$ ) and *Crp4*<sup>+</sup> Paneth cells ( $n = 3$ ) by *in situ* hybridization in proximal jejunal crypts. **c**, BrdU incorporation in ISCs (crypt base columnar cells adjacent to Paneth cells) and progenitors (transit-amplifying cells not adjacent to Paneth cells) after a 4-h pulse ( $n = 4$ ). **d**, Organoid per crypt from vehicle- and GW501516-treated mice ( $n = 3$ ). **e**, Frequencies of flow-sorted ISCs (Lgr5-GFP<sup>hi</sup>) and progenitors (Lgr5-GFP<sup>low</sup>) ( $n = 5$ ) from the entire small intestine of vehicle and GW501516-treated mice. **f**, Organoid-initiating capacity of ISCs derived from vehicle and GW501516-treated mice. Representative images: day-12 organoids ( $n = 5$ ). **g**, Frequency of ISCs (Lgr5-GFP<sup>hi</sup>) in organoids after 14 days of *ex vivo* GW501516

exposure ( $n = 3$ ). **h–j**, Primary (**h**,  $n = 5$ ) and secondary (**i**, **j**,  $n = 5$ ; normalized to vehicle) organoid-forming capacity of control (*Ppard*<sup>L/L</sup> (L, *loxP*) and *Ppard* intestinal knockout (IKO) mice upon *ex vivo* treatment with vehicle, palmitic acid, lipid mixture (L) and GW501516. Representative images: day-4 secondary organoids (**j**). **k**, Normalized clonogenicity of human-derived intestinal organoids after *ex vivo* treatment with palmitic acid, lipid mixture and GW501516 in secondary culture ( $n = 4$ , see Methods). Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (Student's *t*-tests). Scale bars, 20  $\mu$ m (**b**, **c**), 200  $\mu$ m (**f**) and 100  $\mu$ m (**j**). Histological analysis: **b**, *Olfm4*: 15 crypts per group; **c**, *Crp4*: 50 crypts per group; **e**, 50 crypts per group in each experiment. For western blot source data, see Supplementary Fig. 1.

### PPAR- $\delta$ permits non-ISCs to beget tumours

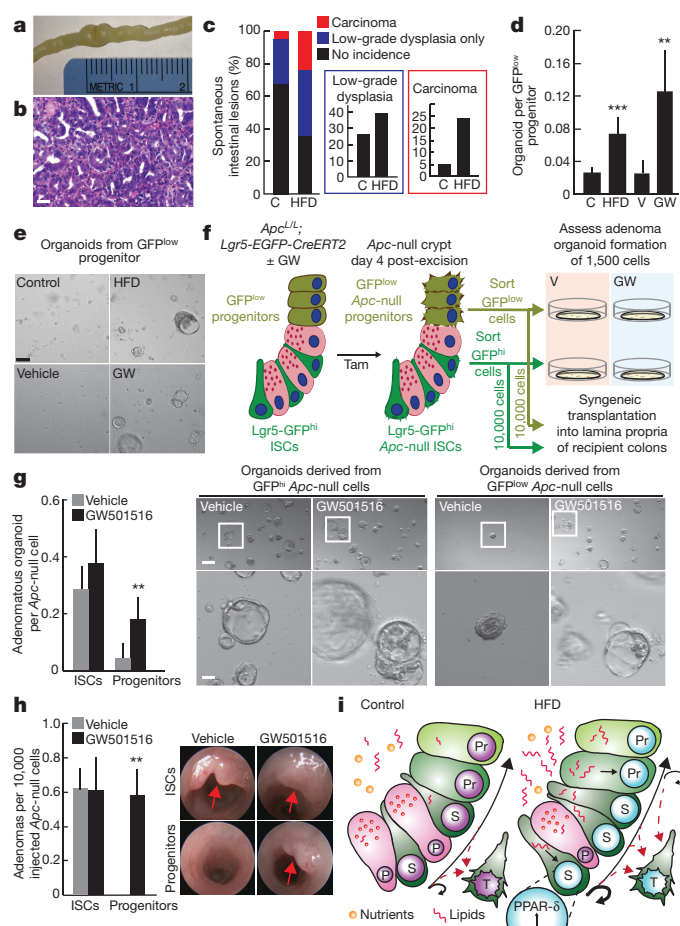
Somatic stem cells often accumulate the initial mutations that lead to oncogenic transformation<sup>3,8,25–27</sup>. We found that in a HFD there is a greater incidence of spontaneous intestinal low-grade dysplastic lesions

(adenomas), carcinomas, or both than in controls (Fig. 5a–c), which may reflect the fact that there are more ISCs in mice on a HFD that can acquire oncogenic mutations. Because HFD-induced PPAR- $\delta$  also activates a  $\beta$ -catenin signature in progenitor cells, we proposed that



**Figure 4 | HFD-induced PPAR- $\delta$  signalling induces expression of a subset of  $\beta$ -catenin target genes.** **a**, **b**, Organoid per crypt quantification from HFD-fed (**a**) and GW501516-treated (**b**) mice with indicated concentrations of Wnt3a. Representative images: day-4 organoids ( $n = 5$ ). **c–f**, Violin plots for most induced  $\beta$ -catenin target genes in ISCs (Lgr5-GFP<sup>hi</sup>) from HFD-fed (**c**) and GW501516-treated (**d**) mice (24 single cells per group), and in progenitors (Lgr5-GFP<sup>low</sup>) from HFD-fed (**e**) and GW501516-treated (**f**) mice (72 single cells per group; see Methods). Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (Student's *t*-tests). Scale bars, 100  $\mu$ m.





**Figure 5 | PPAR- $\delta$  activation confers organoid and tumour-initiating capacity to non-stem cells.** **a**, **b**, Representative spontaneous intestinal tumour from a HFD mouse: gross image (**a**) and microscopic haematoxylin and eosin (H&E) image (**b**). **c**, Incidence of spontaneous intestinal low-grade dysplastic lesions (adenoma) and carcinomas in control ( $n = 19$ ) and HFD ( $n = 25$ ) mice. **d**, **e**, Organoid-initiating capacity of progenitors (Lgr5-GFP<sup>low</sup>) from HFD ( $n = 7$ ) and GW501516-treated ( $n = 5$ ) mice. Representative images: day-7 organoids (**e**). **f**, Schematic assessing *in vitro* and *in vivo* adenoma-initiating capacity of *Apc*-null ISCs (Lgr5-GFP<sup>hi</sup>) and progenitors (Lgr5-GFP<sup>low</sup>) from vehicle- and GW501516-treated mice. Tam, tamoxifen. **g**, Numbers and representative day-5 images of adenomatous organoids from *Apc*-null ISCs (Lgr5-GFP<sup>hi</sup>) and progenitors (Lgr5-GFP<sup>low</sup>) treated with/without GW501516 in EN media (EGF and Noggin only) ( $n = 6$ ). **h**, Optical colonoscopy of tumours formed after orthotopic transplantation of 10,000 *Apc*-null ISCs (Lgr5-GFP<sup>hi</sup>) or *Apc*-null progenitors (Lgr5-GFP<sup>low</sup>) from vehicle- and GW501516-treated mice ( $n = 5$ ). **i**, Model of intestinal adaptation to HFD: mechanistically, HFD activates a PPAR- $\delta$ -mediated program that augments the organoid- and tumour-initiating capacity of intestinal progenitors. A feature of the PPAR- $\delta$  program includes induction of a subset of  $\beta$ -catenin target genes. P, Paneth cell; Pr, progenitor cell; S, stem cell, T, tumour cell. Red dotted lines denote *Apc*-null cells with tumour-forming capability. Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (Student's  $t$ -tests). Scale bars, 50  $\mu$ m (**b**, **g**, bottom) and 200  $\mu$ m (**e**, **g**, top).

non-ISC progenitor populations can acquire stem-cell features and contribute to tumour initiation in diet-induced obesity<sup>3,27</sup>. To explore this possibility, we asked whether HFD or agonist-activated PPAR- $\delta$  influenced the organoid-initiating capacity of progenitors (non-ISCs). Interestingly, Lgr5-GFP<sup>low</sup> progenitors, but not terminally differentiated villous enterocytes, from HFD-fed and PPAR- $\delta$ -agonist-treated mice formed organoids (Fig. 5d, e and Extended Data Figs 4g and 7a, b), raising the possibility that enforced PPAR- $\delta$  signalling in intestinal

progenitors not only bestows organoid-initiating capacity but also tumour-initiating potential.

To test this possibility, we generated *Apc*<sup>L/L</sup>; *Lgr5*-EGFP-IRES-*CreERT2* mice to assess whether pharmacological PPAR- $\delta$  activation modifies the tumorigenic capacity of the Lgr5-GFP<sup>low</sup> progenitors. Injection with tamoxifen leads to *Apc* loss in the Lgr5-GFP<sup>hi</sup> ISCs, which in turn generates *Apc*-null Lgr5-GFP<sup>low</sup> progenitors (Extended Data Fig. 10c). Four days after tamoxifen administration, we isolated *Apc*-null Lgr5-GFP<sup>hi</sup> ISCs and Lgr5-GFP<sup>low</sup> progenitors from vehicle and PPAR- $\delta$ -agonist-treated mice to assess the tumour-forming potential of these populations using separate assays (Fig. 5f, schematic): first, we examined their capacity to give rise to adenomatous organoids in culture; and second, we interrogated the ability of 10,000 sorted stem and progenitor cells to initiate adenomas in syngeneic recipient colons.

We found that *Apc*-null ISCs from PPAR- $\delta$  agonist-treated mice were as clonogenic as those from vehicle controls; however, enforced PPAR- $\delta$  signalling in *Apc*-null progenitors markedly boosted their ability to form adenomatous organoids (Fig. 5g). Next, we assessed the potential of freshly isolated ISC and progenitor cells to form *in vivo* intestinal adenomas (Fig. 5f, schematic). As in the organoid assay, the PPAR- $\delta$  agonist had no additive effect on the ability of *Apc*-null stem cells to form  $\beta$ -catenin<sup>+</sup> (*Apc*-null) adenomas (Fig. 5h and Extended Data Fig. 10a, d) compared to vehicle controls. However, enforced PPAR- $\delta$  signalling permitted *Apc*-null progenitors, but not their vehicle-treated counterparts, to initiate  $\beta$ -catenin<sup>+</sup> (*Apc*-null) adenomas robustly after transplantation into recipient colons (Fig. 5h and Extended Data Fig. 10b, d). These data indicate that PPAR- $\delta$  activation enables a subset of non-ISC progenitors to initiate adenomatous growth *in vitro* and *in vivo* (see Supplementary Information).

## Discussion

Our data favour a model in which a HFD augments ISC self-renewal and bestows features of stemness (that is, organoid-initiating capacity) on non-stem-cell progenitors by activating PPAR- $\delta$  signalling (Fig. 5i). A previous study shows that a different dietary regimen, calorie restriction, increases both stem and Paneth cell numbers and regulates ISC function non-cell autonomously through the Paneth cell niche, with no significant effect on progenitor cell function<sup>1</sup>. By contrast, here we find that a long-term HFD has opposing effects on stem and Paneth cell numbers, and that these stem cells are less dependent on Paneth cells in functional assays. The fact that we find induction of  $\beta$ -catenin targets *Jag1* and *Jag2* (Notch ligands typically elaborated by Paneth cells) in HFD stem and progenitor cells suggests a possible role for Notch signalling. In a HFD, proximate ISCs or progenitor cells may serve as a surrogate source of Notch ligands for Lgr5<sup>+</sup> ISCs not in direct contact with Paneth cells, enabling them to persist *in vivo* and in the organoid assay. A recent report<sup>28</sup> that PPAR- $\delta$  activation in the bone amplifies  $\beta$ -catenin signalling is consistent with our finding that diet-activated PPAR- $\delta$  engages a restricted  $\beta$ -catenin program. Notably, genes within the PPAR- $\delta$ -activated  $\beta$ -catenin program include *Jag1*, *Jag2* and *Bmp4*, which are often deregulated in early intestinal tumorigenesis<sup>29–31</sup> (Fig. 4c–f).

Recent studies propose that intrinsic and extrinsic factors contribute to cancer risk through the control of stem-cell divisions<sup>25,32</sup>. These models predict that extrinsic factors such as a HFD may raise cancer risk by increasing stem-cell divisions, which are the implicated cell-of-origin for many cancers. Our data (Fig. 5i) and a previous study<sup>33</sup> demonstrate that a HFD augments the numbers and proliferation of ISCs, which may partially account for the increase of intestinal tumours in this model of obesity. Another possibility raised by our results is that a HFD-driven PPAR- $\delta$  program also enhances the susceptibility of non-ISCs to undergo oncogenic transformation, thus establishing a larger and more diverse pool of cells capable of initiating tumours. Consistent with this notion, it has been proposed that differentiated cells (non-ISCs) in the background of *Apc*-deficiency

with concurrent activation of oncogenic KRAS and pro-inflammatory NF- $\kappa$ B signalling have the capacity to initiate tumours<sup>9</sup>. Whether and how obesity-related inflammation in a HFD contributes to PPAR- $\delta$  signalling and intestinal tumorigenesis is unknown. In our models, we find no evidence that a HFD or its predominant fatty acid constituents activate inflammatory pathways in intestinal crypts<sup>34</sup> or organoids, respectively (Extended Data Fig. 3a–e).

While some previous work indicates that PPAR- $\delta$  inhibition may have modest anti-cancer effects<sup>19</sup>, sustained PPAR- $\delta$  activation and a HFD have been linked to colorectal cancer initiation and progression<sup>7,35–40</sup>. Future studies will need to address whether PPAR- $\delta$  inhibition in the setting of a HFD affects tumour initiation and progression. Lastly, it will be important to explore if lean ketogenic diets, which like a HFD are composed largely of fatty acids but with fewer carbohydrates, mimic the pro-regenerative effects of a HFD while minimizing the untoward sequelae of obesity.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 22 March 2015; accepted 28 January 2016.**

- Yilmaz, Ö. H. *et al.* mTORC1 in the Paneth cell niche couples intestinal stem-cell function to calorie intake. *Nature* **486**, 490–495 (2012).
- Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
- Mihaylova, M. M., Sabatini, D. M. & Yilmaz, Ö. H. Dietary and metabolic control of stem cell function in physiology and cancer. *Cell Stem Cell* **14**, 292–305 (2014).
- Sato, T. *et al.* Paneth cells constitute the niche for *Lgr5* stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
- Finucane, M. M. *et al.* National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet* **377**, 557–567 (2011).
- Calle, E. E. & Kaaks, R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nature Rev. Cancer* **4**, 579–591 (2004).
- Baltgalvis, K. A., Berger, F. G., Pena, M. M., Davis, J. M. & Carson, J. A. The interaction of a high-fat diet and regular moderate intensity exercise on intestinal polyp development in *Apc*<sup>Min/+</sup> mice. *Cancer Prev. Res. (Phila.)* **2**, 641–649 (2009).
- Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009).
- Schwitalla, S. *et al.* Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* **152**, 25–38 (2013).
- Eckel-Mahan, K. L. *et al.* Reprogramming of the circadian clock by nutritional challenge. *Cell* **155**, 1464–1478 (2013).
- Winzell, M. S. & Ahren, P. The high-fat diet-fed mouse: a model for studying mechanisms and treatment of impaired glucose tolerance and type 2 diabetes. *Diabetes* **53** (suppl. 3), S215–S219 (2004).
- Schuijers, J., van der Flier, L. G., van Es, J. & Clevers, H. Robust Cre-mediated recombination in small intestinal stem cells utilizing the *Olfm4* locus. *Stem Cell Reports* **3**, 234–241 (2014).
- Sato, T. *et al.* Single *Lgr5* stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
- Schuijers, J. *et al.* *Ascl2* acts as an R-spondin/Wnt-responsive switch to control stemness in intestinal crypts. *Cell Stem Cell* **16**, 158–170 (2015).
- Marsh, V. *et al.* Epithelial Pten is dispensable for intestinal homeostasis but suppresses adenoma development and progression after *Apc* mutation. *Nature Genet.* **40**, 1436–1444 (2008).
- Buettner, R. *et al.* Defining high-fat-diet rat models: metabolic and molecular effects of different fat types. *J. Mol. Endocrinol.* **36**, 485–501 (2006).
- Peters, J. M., Shah, Y. M. & Gonzalez, F. J. The role of peroxisome proliferator-activated receptors in carcinogenesis and chemoprevention. *Nature Rev. Cancer* **12**, 181–195 (2012).
- Tontonoz, P. & Spiegelman, B. M. Fat and beyond: the diverse biology of PPAR- $\gamma$ . *Annu. Rev. Biochem.* **77**, 289–312 (2008).
- Neels, J. G. & Grimaldi, P. A. Physiological functions of peroxisome proliferator-activated receptor  $\beta$ . *Physiol. Rev.* **94**, 795–858 (2014).
- Ito, K. *et al.* A PML-PPAR- $\delta$  pathway for fatty acid oxidation regulates hematopoietic stem cell maintenance. *Nature Med.* **18**, 1350–1358 (2012).
- Narkar, V. A. *et al.* AMPK and PPAR- $\delta$  agonists are exercise mimetics. *Cell* **134**, 405–415 (2008).
- van der Flier, L. G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu. Rev. Physiol.* **71**, 241–260 (2009).
- van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903–912 (2009).
- Muñoz, J. *et al.* The *Lgr5* intestinal stem cell signature: robust expression of proposed quiescent ‘+4’ cell markers. *EMBO J.* **31**, 3079–3091 (2012).
- Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
- Yilmaz, Ö. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475–482 (2006).
- Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337 (2013).
- Schlotysek, C. *et al.* PPAR- $\delta$  governs Wnt signaling and bone turnover. *Nature Med.* **19**, 608–613 (2013).
- Rodilla, V. *et al.* Jagged1 is the pathological link between Wnt and Notch pathways in colorectal cancer. *Proc. Natl Acad. Sci. USA* **106**, 6315–6320 (2009).
- Kumar, S. R. *et al.* Preferential induction of EphB4 over EphB2 and its implication in colorectal cancer progression. *Cancer Res.* **69**, 3736–3745 (2009).
- Kim, J. S. *et al.* Oncogenic  $\beta$ -catenin is required for bone morphogenetic protein 4 expression in human cancer cells. *Cancer Res.* **62**, 2744–2748 (2002).
- Wu, S., Powers, S., Zhu, W. & Hannun, Y. A. Substantial contribution of extrinsic risk factors to cancer development. *Nature* **529**, 43–47 (2016).
- Mah, A. T., Van Landeghem, L., Gavin, H. E., Magness, S. T. & Lund, P. K. Impact of diet-induced obesity on intestinal stem cells: hyperproliferation but impaired intrinsic function that requires insulin/IGF1. *Endocrinology* **155**, 3302–3314 (2014).
- Johnson, A. M. *et al.* High fat diet causes depletion of intestinal eosinophils associated with intestinal permeability. *PLoS ONE* **10**, e0122195 (2015).
- Wang, D. *et al.* Peroxisome proliferator-activated receptor  $\delta$  promotes colonic inflammation and tumor growth. *Proc. Natl Acad. Sci. USA* **111**, 7084–7089 (2014).
- Wang, D. *et al.* Crosstalk between peroxisome proliferator-activated receptor  $\delta$  and VEGF stimulates cancer progression. *Proc. Natl Acad. Sci. USA* **103**, 19069–19074 (2006).
- Park, B. H., Vogelstein, B. & Kinzler, K. W. Genetic disruption of PPAR- $\delta$  decreases the tumorigenicity of human colon cancer cells. *Proc. Natl Acad. Sci. USA* **98**, 2598–2603 (2001).
- Zuo, X. *et al.* Targeted genetic disruption of peroxisome proliferator-activated receptor- $\delta$  and colonic tumorigenesis. *J. Natl. Cancer Inst.* **101**, 762–767 (2009).
- Gupta, R. A. *et al.* Activation of nuclear hormone receptor peroxisome proliferator-activated receptor- $\delta$  accelerates intestinal adenoma growth. *Nature Med.* **10**, 245–247 (2004).
- Barak, Y. *et al.* Effects of peroxisome proliferator-activated receptor  $\delta$  on placenta, adiposity, and colorectal cancer. *Proc. Natl Acad. Sci. USA* **99**, 303–308 (2002).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by the Howard Hughes Medical Institute (S.H.O. and D.M.S.), Ellison Medical Foundation Aging grant (D.M.S.), NIH (R01 CA103866 and AI47389; D.M.S.), NIH (K08 CA198002; J.R.), Department of Defense PRCRP Career Development Award CA120198 (J.R.), NIH (R00 AG045144; Ö.H.Y.), NIH (R00 AG041765; D.W.L.), Center for the Study of Inflammatory Bowel Diseases from the Massachusetts General Hospital NIH (DK043351; Ö.H.Y.), NIH Cancer Center Support (core) grant P30-CA14051 (Ö.H.Y.), Kathy and Curt Marble Cancer Research Fund (Ö.H.Y.), American Federation of Aging Research (AFAR; Ö.H.Y.), and V Foundation Scholar grant (J.R. and Ö.H.Y.). M.D.M. is supported by a Koch MIT Ludwig Center post-doctoral fellowship, D.K. receives fellowship support from MGH (T32DK007191), and M.M.M. is a Robert Black Fellow of the Damon Runyon Cancer Research Foundation. We thank the Koch Institute Swanson Biotechnology Center (SBC) for technical support, specifically the Hope Babette Tang (1983) Histology Facility and Kathleen Cormier. We thank S. Holder for superior histology and help with special stains. We thank P. Wisniewski and G. Paradis of the Whitehead flow cytometry and Koch core facilities, respectively, for their expertise in cell sorting. We thank members of the Yilmaz laboratory for discussions.

**Author Contributions** Ö.H.Y., S.B., M.D.M. and J.R. performed all experiments, and participated in their design and interpretation. J.R. optimized the colonoscopy transplantation assay, with help from A.A., M.D.M. and Ö.H.Y. S.-J.H. performed the mRNA-sequencing and analysis, with help from G.W.B., S.B., L.P. and Y.K. K.E.B.-R., A.A. and M.D.M. performed and interpreted the immunohistochemistry and *in situ* hybridization, under the guidance of Ö.H.Y. S.B. performed the single-cell analysis, with assistance from M.E.X., R.D., G.G., G.-C.Y. and A.S. M.S. and G.P.N. performed electron microscopy and helped with its interpretation. S.S., V.D. and Ö.H.Y. performed all pathology on the mice with help from J.R., and participated in the design and interpretation of experiments. M.E.X., E.A., R.D., M.A.-R., M.M.M. and D.W.L. supplied HFD-fed mice and provided experimental support. C.R.F. and N.G. provided assistance with the acquisition of human samples, while D.K. performed and interpreted the human cell culture experiments. S.H.O. and D.M.S. participated in the design and interpretation of experiments. Ö.H.Y. wrote the paper with support from S.B., M.D.M. and J.R.

**Author Information** RNA-sequencing data have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE67324. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.M.S. ([sabatini@wi.mit.edu](mailto:sabatini@wi.mit.edu)) or Ö.H.Y. ([ohyilmaz@mit.edu](mailto:ohyilmaz@mit.edu)).



## METHODS

**Mice, HFD and drug treatment.** Mice were housed in the Unit for Laboratory Animal Medicine at the Whitehead Institute for Biomedical Research and Koch Institute for Integrative Cancer Research. The following strains were obtained from the Jackson Laboratory: *Lgr5-EGFP-IRES-CreERT2* (strain name: B6.129P2-Lgr5<sup>tm1(cre/ERT2)Cle</sup>/J, stock number 008875), *Rosa26-lacZ* (strain name: B6.129S4-Gt(ROSA)26Sor<sup>tm1Sor</sup>/J, stock number 003474), *db/db* (strain name: B6.BKS(D)-Lepr<sup>ib</sup>/J, stock number 000697), *Ppard<sup>L/L</sup>* (strain name: B6.129S4-Ppard<sup>tm1Rev</sup>/J, stock number 005897). *Apc<sup>loxP exon 14</sup>* (*Apc<sup>L/L</sup>*) has been previously described<sup>41</sup>. *Villin-CreERT2* was a gift from S. Robine. Long-term HFD was achieved by feeding male and female mice a dietary chow consisting of 60% kcal fat (Research Diets D12492) beginning at the age of 8–12 weeks and extending for a period of 9–14 months. Control mice were sex- and age-matched and fed standard chow *ad libitum*. GW501516 (Enzo) was reconstituted in DMSO at 4.5 mg ml<sup>-1</sup> and diluted 1:10 in a solution of 5% PEG400 (Hampton Research), 5% Tween80 (Sigma), 90% H<sub>2</sub>O for a daily intraperitoneal injection of 4 mg kg<sup>-1</sup>. *Apc* exon 14 was excised by tamoxifen suspended in sunflower seed oil (Spectrum S1929) at a concentration of 10 mg ml<sup>-1</sup> and 250 µl per 25 g of body weight, and administered by intraperitoneal injection twice over 4 days before collecting tissue. *Ppard<sup>L/L</sup>* mice were administered 4–5 intraperitoneal injections of tamoxifen on alternate days. Mice were analysed within 2 weeks of the last tamoxifen injection. BrdU was prepared at 10 mg ml<sup>-1</sup> in PBS, passed through a 0.22-µm filter and injected at 100 mg kg<sup>-1</sup>.

**Immunohistochemistry and immunofluorescence.** As previously described<sup>41</sup>, tissues were fixed in 10% formalin, paraffin embedded and sectioned. Antigen retrieval was performed with Borg Decloaker RTU solution (Biocare Medical) in a pressurized Decloaking Chamber (Biocare Medical) for 3 min. Antibodies used: rat anti-BrdU (1:2,000 (immunohistochemistry (IHC)), 1:1,000 (immunofluorescence (IF)) Abcam 6326), rabbit chromogranin A (1:4,000 (IHC), 1:250 (IF), Abcam 15160), rabbit monoclonal non-phospho  $\beta$ -catenin (1:800 (IHC), 1:400 (IF), CST 8814S), mouse monoclonal  $\beta$ -catenin (1:200, BD Biosciences 610154), rabbit polyclonal lysozyme (1:250, Thermo RB-372-A1), rabbit polyclonal MUC2 (1:100, Santa Cruz Biotechnology 15334), rabbit monoclonal OLFM4 (1:10,000, gift from CST, clone PP7), Biotin-conjugated secondary donkey anti-rabbit or anti-rat antibodies were used from Jackson ImmunoResearch. The Vectastain Elite ABC immunoperoxidase detection kit (Vector Labs PK-6101) followed by Dako Liquid DAB+ Substrate (Dako) was used for visualization. For immunofluorescence, Alexa Fluor 568 secondary antibody (Invitrogen) was used with Prolong Gold (Life Technologies) mounting media. All antibody incubations involving tissue or sorted cells were performed with Common Antibody Diluent (Biogenex). Organoids were fixed with 4% paraformaldehyde, permeabilized with 0.5% Triton X-100 in PBS, rinsed with 100 mM glycine in PBS, blocked with 10% donkey serum in PBS, incubated overnight with primary antibody at 4 °C, rinsed and incubated with Alexa Fluor 568 secondary antibody (Invitrogen), and mounted with Prolong Gold (Life Technologies) mounting media.

**In situ hybridization.** The *in situ* hybridization probes used in this study correspond to expressed sequence tags or fully sequenced cDNAs obtained from Open Biosystems. The accession numbers (IMAGE mouse cDNA clone in parenthesis) for these probes are as follows: mouse *Olfm4* BC141127 (9055739), mouse *Crp4* BC134360 (40134597). Both sense and antisense probes were generated to ensure specificity by *in vitro* transcription using DIG RNA labelling mix (Roche) according to the manufacturer's instructions and to previously published detailed methods<sup>23,42</sup>. Single-molecule *in situ* hybridization was performed using Advanced Cell Diagnostics RNAscope 2.0 HD Detection Kit.

**Radiation and clonogenic microcolony assay.** Adult mice were exposed to 15 Gy of ionizing irradiation from a 137-caesium source (GammaCell) and euthanized after 72 h. The number of surviving crypts per length of the intestine was enumerated from haematoxylin-and-eosin-stained sections<sup>15</sup>.

**Immunoprecipitation and immunoblotting.** Antibodies: rabbit polyclonal anti-PPAR- $\delta$  (1:100, Thermo PA1-823A), rabbit polyclonal anti-CPT1a (1:250, ProteinTech 15184-1-AP), rabbit polyclonal anti-HMGCS2 (1:500, Sigma AV41562), rabbit monoclonal anti-FABP1 (1:1,000, Abcam ab129203), NF- $\kappa$ B Sampler Pathway Kit (CST, 9936S), mouse monoclonal anti-STAT-3 (CST, 9139P), rabbit monoclonal anti-P-STAT3 (Y705) XP (CST, 9145P), mouse monoclonal anti-CREB (CST, 86B10), mouse monoclonal anti- $\beta$ -catenin (1:200, BD Biosciences 610154), rabbit polyclonal anti- $\gamma$ -tubulin (1:1,000, Sigma T5192). For immunoprecipitation assays, crypts were collected and nuclear extraction was carried out using Abcam nuclear extraction kit (ab113474) following manufacturer's instructions. Nuclear extracts were incubated with 5 µg anti-PPAR- $\delta$  antibody (Thermo), or anti-rabbit IgG control antibody (Santa Cruz) overnight at 4 °C followed by 2 h of incubation with Dynabeads Protein G for immunoprecipitation. Protein

complexes bound to antibody and beads were washed five times and eluted with Laemmli sample buffer. Samples were resolved by SDS–PAGE. Protein interaction was analysed by immunoblotting.

Lgr5-GFP<sup>hi</sup> ISCs or Lgr5-GFP<sup>low</sup> progenitors were sorted directly into Laemmli sample buffer and boiled for 5 min. Samples were resolved by SDS–PAGE and analysed by immunoblotting with horseradish peroxidase (HRP)-conjugated IgG secondary antibodies (1:10,000, Santa Cruz Biotechnology sc-2054) and Western Lightning Plus-ECL detection kit (Perkin Elmer NEL104001EA).

**Flow cytometry and isolation of ISCs, colonic stem cells and Paneth cells.** As previously reported and briefly summarized here, small intestines and colons were removed, washed with cold PBS without magnesium chloride and calcium (PBS–/–) opened longitudinally, and then cut into 3–5-mm fragments. Pieces were washed several times with cold PBS–/– until clean, washed 2–3 with PBS–/– EDTA (10 mM), incubated on ice for 90–120 min, and gently shook at 30-min intervals. Crypts were then mechanically separated from the connective tissue by more rigorous shaking, and then filtered through a 70-µm mesh into a 50-ml conical tube to remove villus material (for small intestine) and tissue fragments. Crypts were removed from this step for crypt culture experiments and embedded in Matrigel with crypt culture media. For ISC isolation, the crypt suspensions were dissociated to individual cells with TrypLE Express (Invitrogen). Cell labelling consisted of an antibody cocktail comprising CD45-PE (eBioscience, 30-F11), CD31-PE (Biolegend, Mec13.3), Ter119-PE (Biolegend, Ter119), CD24-Pacific Blue (Biolegend, M1/69), CD117-APC/Cy7 (Biolegend, 2BS), and EPCAM-APC (eBioscience, G8.8). ISCs were isolated as Lgr5-EGFP<sup>hi</sup>Epcam<sup>+</sup>CD24<sup>low</sup>–CD31<sup>–</sup>Ter119<sup>–</sup>CD45<sup>–</sup>7-AAD<sup>–</sup>. EGFP<sup>low</sup> progenitors were isolated as EGFP<sup>low</sup>Epcam<sup>+</sup>CD24<sup>low</sup>–CD31<sup>–</sup>Ter119<sup>–</sup>CD45<sup>–</sup>7-AAD<sup>–</sup>, and Paneth cells from small intestine were isolated as CD24<sup>hi</sup>Sidescatter<sup>hi</sup>Lgr5-EGFP<sup>–</sup>Epcam<sup>+</sup>CD31<sup>–</sup>Ter119<sup>–</sup>CD45<sup>–</sup>7-AAD<sup>–</sup> with a BD FACS Aria II SORP cell sorter into supplemented crypt culture medium for culture. Dead cells were excluded from the analysis with the viability dye 7-AAD (Life Technologies). When indicated, populations were cytopun (Thermo Cytospin 4) at 800 r.p.m. for 2 min, or allowed to settle at 37 °C in fully humidified chambers containing 5% CO<sub>2</sub> onto poly-L-lysine-coated slides (Polysciences). The cells were subsequently fixed in 4% paraformaldehyde (pH 7.4, Electron Microscopy Sciences) before staining.

**Culture media for crypts and isolated cells.** Isolated crypts were counted and embedded in Matrigel (Corning 356231 growth factor reduced) at 5–10 crypts per µl and cultured in a modified form of medium as described previously<sup>13</sup>. Unless otherwise noted, Advanced DMEM (Gibco) was supplemented by EGF 40 ng ml<sup>-1</sup> (R&D), Noggin 200 ng ml<sup>-1</sup> (Peprotech), R-spondin 500 ng ml<sup>-1</sup> (R&D or Sino Biological), N-acetyl-L-cysteine 1 µM (Sigma-Aldrich), N2 1X (Life Technologies), B27 1X (Life Technologies), Chiron 10 µM (Stemgent), Y-27632 dihydrochloride monohydrate 20 ng ml<sup>-1</sup> (Sigma-Aldrich). Colonic crypts were cultured in 50% conditioned medium derived from L-WRN cells supplemented with Y-27632 dihydrochloride monohydrate 20 ng ml<sup>-1</sup>, as described<sup>43</sup>. Approximately 25–30 µl droplets of Matrigel with crypts were plated onto a flat bottom 48-well plate (Corning 3548) and allowed to solidify for 20–30 min in a 37 °C incubator. Three hundred microlitres of crypt culture medium was then overlaid onto the Matrigel, changed every 3 days, and maintained at 37 °C in fully humidified chambers containing 5% CO<sub>2</sub>. Clonogenicity (colony-forming efficiency) was calculated by plating 50–300 crypts and assessing organoid formation 3–7 days or as specified after initiation of cultures. Palmitic acid (Cayman Chemical Company 10006627 conjugated to BSA), oleic acid (Sigma O1008), lipid mixture (Sigma L0288), or GW501516 (Enzo) were added immediately to cultures at 30 µM (palmitic acid, oleic acid), 2% (lipid mixture), and 1 µM (GW501516). 4-OH tamoxifen (Calbiochem, 579002, 10 nM) was added to organoid cultures derived from *Ppard<sup>L/L</sup>*; *Villin-CreERT2* (*Ppard* IKO) crypts to ensure *Ppard* excision in the *ex vivo* fatty acid or GW501516 experiments.

Isolated ISCs or progenitor cells were centrifuged for 5 min at 250g, resuspended in the appropriate volume of crypt culture medium (500–1,000 cells µl<sup>-1</sup>), then seeded onto 25–30 µl Matrigel (Corning 356231 growth factor reduced) containing 1 µM Jagged (Ana-Spec) in a flat bottom 48-well plate (Corning 3548). Alternatively, ISCs and Paneth cells were mixed after sorting in a 1:1 ratio, centrifuged, and then seeded onto Matrigel. The Matrigel and cells were allowed to solidify before adding 300 µl of crypt culture medium. The crypt media was changed every second or third day. Organoids were quantified on days 3, 7 and 10 of culture, unless otherwise specified.

**Secondary mouse organoid assays.** For secondary organoid assays, either individual primary organoids or many primary organoids were mechanically dissociated and then replated, or organoids were dissociated for 10 min in TrypLE Express at 32 °C, resuspended with SMEM (Life Technologies), centrifuged (5 min at 250g) and then resuspended in cold SMEM with the viability dye 7-AAD. Live cells were sorted and seeded onto Matrigel as previously described in standard crypt



media (not supplemented with lipids or GW501516). Secondary organoids were enumerated on day 4, unless otherwise specified.

**Human crypt cultures.** Human biopsies were obtained from patients with informed consent undergoing intestinal resection at the Massachusetts General Hospital (MGH). The MGH Institutional Review Board committee and Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects approved the study protocols. Crypts were isolated<sup>43</sup>, embedded in Matrigel and subsequently exposed to lipid mixture, palmitic acid or GW501516 (as described in earlier). Cultures were passaged weekly and maintained for 3–4 weeks. To passage, equal numbers of organoids from each condition were disrupted with trypsin/EDTA. Numbers of organoids were counted 4–7 days after passaging into control media. Counts were normalized to numbers of organoids present in control wells and plotted. Statistical significance was calculated by performing analysis of variance (ANOVA) multiple comparisons of the means for each group. For quantitative RNA expression analysis, organoids were dissociated, cells were selected as a live population by flow cytometry (7-AAD, Life Technologies), and sorted into Tri Reagent (Life Technologies) for RNA isolation.

**Electron microscopy.** After 5 days of culturing, intestinal organoids were placed into Karnovsky's KII solution (2.5% glutaraldehyde, 2.0% paraformaldehyde, 0.025% calcium chloride, in a 0.1 M sodium cacodylate buffer, pH 7.4) and fixed overnight. Subsequently, they were post-fixed in 2.0% osmium tetroxide, stained en bloc with uranyl acetate, dehydrated in graded ethanol solutions, infiltrated with propylene oxide/Epon mixtures, flat embedded in pure Epon, and polymerized overnight at 60 °C. Then 1- $\mu$ m sections were cut, stained with toluidine blue, and examined by light microscopy. Representative areas were chosen for electron microscopic study and the Epon blocks were trimmed accordingly. Thin sections were cut with an LKB 8801 ultramicrotome and diamond knife, stained with Sato's lead, and examined in a FEI Morgagni transmission electron microscope. Images were captured with an AMT (Advanced Microscopy Techniques) 2K digital CCD camera.

**RNA isolation.** For RNA sequencing (RNA-seq), total RNA was extracted from 200,000 sorted Lgr5-GFP<sup>hi</sup> ISC and Lgr5-GFP<sup>low</sup> progenitors by pooling 2–5 71-week-old HFD male or control mice using Tri Reagent (Life Technologies) according to the manufacturer's instructions, except for an overnight isopropanol precipitation at –20 °C. From the total RNA, poly(A)<sup>+</sup> RNA was selected using Oligo(dT)<sub>25</sub>-Dynabeads (Life technologies) according to the manufacturer's protocol.

**RNA-seq library preparation.** Strand-specific RNA-seq libraries were prepared using the dUTP-based, Illumina-compatible NEXTFlex Directional RNA-Seq Kit (Bioo Scientific) according to the manufacturer's directions. All libraries were sequenced with an Illumina HiSeq 2000 sequencing machine.

**Processing of RNA-seq reads and measuring expression level.** For RNA-seq data analysis, raw stranded reads (40 nucleotides) were trimmed to remove adaptor and bases with quality scores below 20, and reads shorter than 35 nucleotides were excluded. High-quality reads were mapped to the mouse genome (mm10) with TopHat version 1.4.1 (ref. 44), using known splice junctions from Ensembl Release 70 and allowing at most two mismatches. Genes were quantified with htseq-count (with the 'intersect strict' mode) using Ensembl Release 70 gene models. Gene counts were normalized across all samples using estimateSizeFactors from the DESeq R/Bioconductor package<sup>45</sup>. Differential expression analysis was also performed between two samples of interest with DESeq. GSEA (<http://software.broadinstitute.org/gsea/index.jsp>) was performed by using the pre-ranked (according to their ratios) 8,240 differentially expressed genes as the expression data set. Motif Analysis was performed using Haystack motif enrichment tool: <http://github.com/lucapinello/Haystack><sup>46</sup>.

**Single-cell gene expression analysis.** In total, 24 single Lgr5-GFP<sup>hi</sup> ISCs and 72 single Lgr5-GFP<sup>low</sup> progenitor cells were sorted from control or HFD-fed mice ( $n = 2$  mice per group) for single-cell gene expression analysis. For one-tube single-cell sequence-specific preamplification, individual primer sets of  $\beta$ -catenin target genes (total of 96, Supplementary Table 2) were pooled to a final concentration of 0.1 mM for each primer. Single cells were directly sorted into 96-well plates containing 5  $\mu$ l RT-PCR master mix (2.5  $\mu$ l CellsDirect reaction mix, Invitrogen; 0.5  $\mu$ l primer pool; 0.1  $\mu$ l reverse transcriptase/Taq enzyme, Invitrogen; 1.9  $\mu$ l nuclease-free water) in each well. Immediately after, plates were placed on PCR machine for preamplification. Sequence-specific preamplification PCR protocol was as following: 60 min at 50 °C for cell lysis and sequence-specific reverse transcription; then 3 min at 95 °C for reverse transcriptase inactivation and Taq polymerase activation. cDNA was then amplified by 20 cycles of 15 s at 95 °C for initial denaturation, 15 min at 60 °C for annealing and elongation. After pre-amplification, samples were diluted 1:5 before high-throughput microfluidic real-time PCR analysis using Fluidigm platform. Amplified single-cell cDNA samples were assayed for gene expression using individual qRT-PCR primers

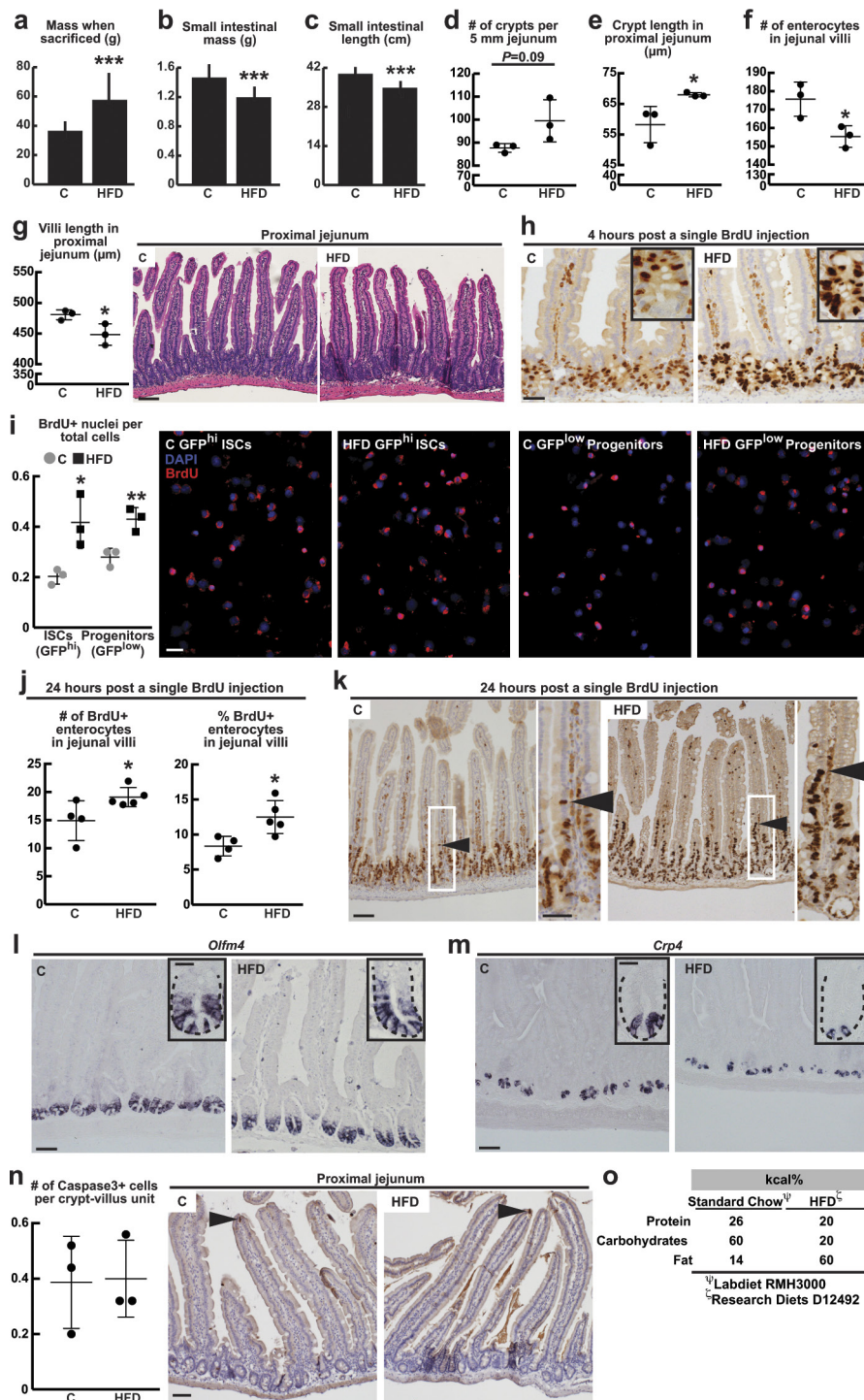
and 96.96 dynamic arrays on a BioMark System by following manufacturers protocol (Fluidigm). To confirm PPAR- $\delta$ -mediated induction of the most upregulated genes ( $n = 3$  mice, 24 ISCs and 72 progenitors per group), or for single-cell analysis of organoid composition ( $n = 3$  mice, 48 cells per group) and *db/db* mice ( $n = 3$ , 48 cells per group) standard single-cell qRT-PCR was performed using preamplified cDNA with corresponding primers. For Fluidigm analysis, threshold cycle ( $C_t$ ) values were calculated using the BioMark Real-Time PCR Analysis software (Fluidigm). See Supplementary Information for raw gene expression data. Gene expression levels were estimated by subtracting the  $C_t$  values from the background level of 35, which approximately represent the log<sub>2</sub> gene expression levels. The t-Distributed stochastic neighbour embedding (t-SNE) analysis<sup>47</sup> was performed using the MATLAB toolbox for dimensionality reduction. Differential expression analysis was conducted using the two-sided Wilcoxon–Mann–Whitney rank sum test implemented in the R coin package (<https://www.r-project.org>). *P* values were adjusted for multiple testing<sup>48</sup> using the *p.adjust* function in R with method = 'fdr' option. Fold changes were calculated as the difference of median of log<sub>2</sub> expression levels for the two cell populations. Split violin plots were generated using the vioplot package and the vioplot2 function in R (<https://gist.github.com/mbjoseph/5852613>). The heatmap for  $\beta$ -catenin target genes was generated with the MultiExperiment Viewer (MeV) program (<http://www.tm4.org/mev.html>) using the correlation-based distance and average linkage method as parameters of the unsupervised hierarchical clustering of genes. The heatmap for organoid composition was generated using MATLAB. The percentages of *Jag1/Jag2*-upregulated cells were calculated based on the number of single cells whose log<sub>2</sub> expression was above 15.

**qRT-PCR.** Approximately 25,000 cells were sorted into Tri Reagent (Life Technologies) and total RNA was isolated according to the manufacturer's instructions with following modification: the aqueous phase containing total RNA was purified using the RNeasy plus kit (Qiagen). RNA was converted to cDNA with the cDNA synthesis kit (Bio-Rad). qRT-PCR was performed with diluted cDNA (1:5) in three wells for each primer and SYBR green master mix (Bio-Rad) on Bio-Rad iCycler RT-PCR detection system. For organoid experiments, 1,000 live cells were sorted and qRT-PCR optimized for low cell numbers (<1,000) was performed after sequence specific pre-amplification (cDNA diluted 1:200 in three wells for each primer) as described in single-cell gene expression analysis. All qRT-PCR experiments were repeated at least three independent times. Primers used are listed on Supplementary Table 1.

**Orthotopic transplantation.** *Apc*<sup>L/L</sup>; *Lgr5-EGFP-IRES-CreERT2* mice were treated with vehicle or GW501516 for 1 month, and then injected with two intraperitoneal doses of tamoxifen. Four days later, *Apc*-null Lgr5-GFP<sup>hi</sup> ISCs and Lgr5-GFP<sup>low</sup> progenitors were sorted by flow cytometry, as described earlier. For primary cell transplantations, 10,000 *Apc*-null Lgr5-GFP<sup>hi</sup> ISCs and Lgr5-GFP<sup>low</sup> progenitors were resuspended into 90% crypt culture media (as described) and 10% Matrigel, then transplanted into the colonic lamina propria of C57BL/6 recipient mice by optical colonoscopy using a custom injection needle (Hamilton Inc., 33-gauge, small Hub RN NDL, 16 inches long, point 4, 45 degree bevel, like part number 7803-05), syringe (Hamilton Inc. part number 7656-01), and transfer needle (Hamilton Inc. part number 7770-02). Optical colonoscopy was performed using a Karl Storz Image 1 HD Camera System, Image 1 HUB CCU, 175 Watt Xenon Light Source, and Richard Wolf 1.9mm/9.5 Fr Integrated Telescope (part number 8626.431). Four injections were performed per mouse. Mice then underwent colonoscopy 8 weeks later to assess tumour formation. Colonoscopy videos and images were saved for offline analysis. Following sacrifice, the distal colons were excised and fixed in 10% formalin, then examined by haematoxylin and eosin section to identify adenomas. Histology images were reviewed by gastrointestinal pathologists who were blinded to the treatment groups (S.S., V.D. and Ö.H.Y.).

**Statistics and animal models.** All experiments reported in Figs 1–5 were repeated at least three independent times, except for Figs 3a, 4c, d, which were repeated twice. All samples represent biological replicates. For mouse organoid assays, 2–4 wells per group with at least 3 different mice were analysed. For human organoid assays, 4 wells per group with 4 different patient samples were analysed and experiments were repeated 4 times. All centre values shown in graphs refer to the mean. For statistical significance of the differences between the means of two groups, we used two-tailed Student's *t*-tests. Statistical significance in Fig. 3k was calculated by performing ANOVA multiple comparisons of the means for each group. No samples or animals were excluded from analysis, and sample size estimates were not used. Animals were randomly assigned to groups. Studies were not conducted blinded, with the exception of all histological analyses and Fig. 5c, h. All experiments involving mice were carried out with approval from the Committee for Animal Care at MIT and under supervision of the Department of Comparative Medicine at MIT.

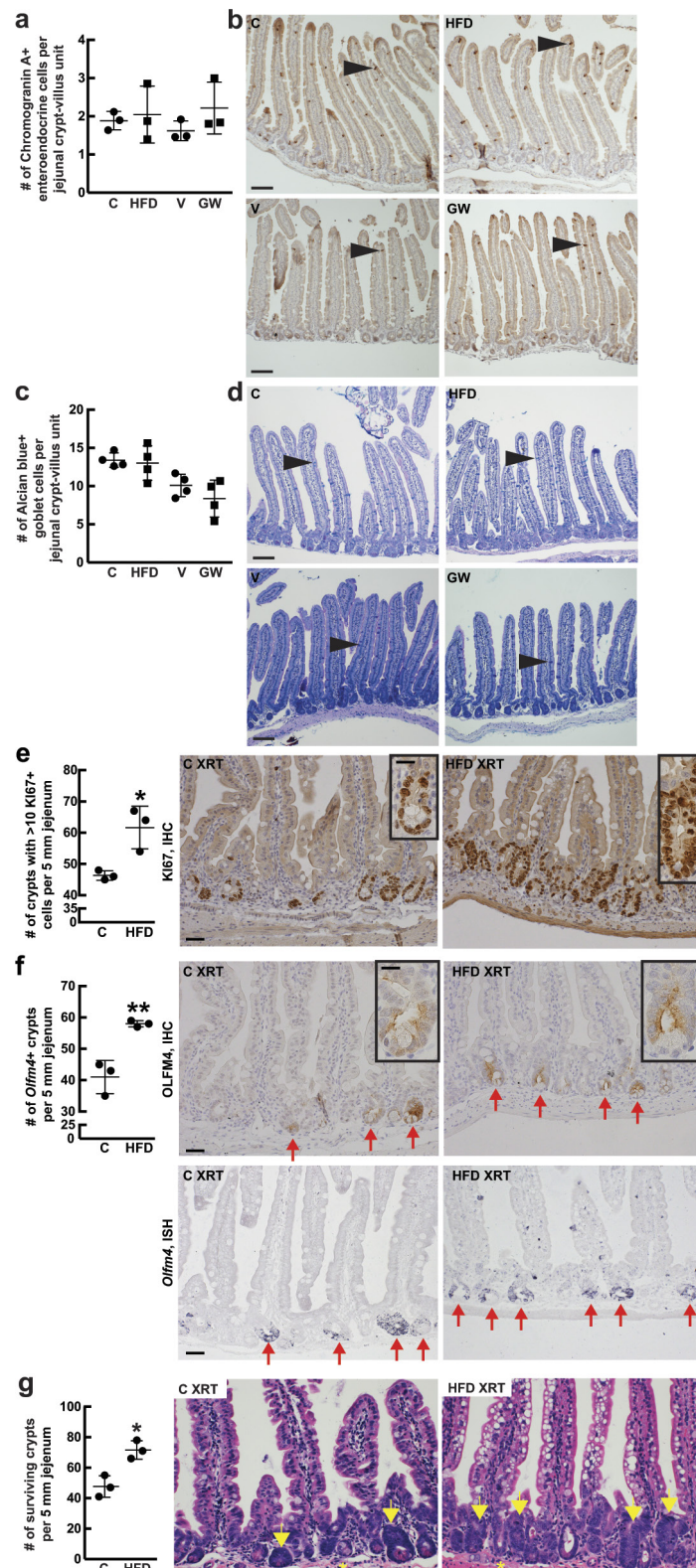
41. Colnot, S. *et al.* Colorectal cancers in a new mouse model of familial adenomatous polyposis: influence of genetic and environmental modifiers. *Lab. Invest.* **84**, 1619–1630 (2004).
42. Gregorieff, A. & Clevers, H. *In situ* hybridization to identify gut stem cells. *Curr. Protoc. Stem Cell Biol.* **Chapter 2**, Unit 2F.1 (2010).
43. Miyoshi, H. & Stappenbeck, T. S. *In vitro* expansion and genetic modification of gastrointestinal stem cells in spheroid culture. *Nature Protocols* **8**, 2471–2482 (2013).
44. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
45. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
46. Pinello, L., Xu, J., Orkin, S. H. & Yuan, G. C. Analysis of chromatin-state plasticity identifies cell-type-specific regulators of H3K27me3 patterns. *Proc. Natl Acad. Sci. USA* **111**, E344–E353 (2014).
47. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
48. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* **57**, 289–300 (1995).



**Extended Data Figure 1 | A HFD alters intestinal morphology and enhances intestinal progenitor proliferation.** **a–g.** In comparison to mice fed a standard chow, mice on a HFD gained on average 50% mass (**a**, control:  $n = 11$ , HFD:  $n = 15$ ), had reduced small intestinal mass and length (**b**, **c**, control:  $n = 11$ , HFD:  $n = 15$ ), longer crypts and shorter villi (**e**, **g**,  $n = 3$  each), and fewer villus enterocytes (**f**,  $n = 3$ ). HFD did not change the density of crypts (**d**,  $n = 3$ ) in the proximal jejunum. The proximal jejunum was defined as the length between 6 and 9 cm as measured from the pylorus (the distal portion of the stomach). **h–k.** HFD enhanced BrdU incorporation in ISCs (or crypt base columnar cells) and progenitor cells (or transit-amplifying cells) in the proximal jejunum (**h**,  $n = 6$ ) and sorted cell populations (**i**,  $n = 3$ ) after a 4-h pulse. HFD increased the total (**j**, control:  $n = 4$ , HFD:  $n = 5$ ) and normalized numbers of BrdU-labelled enterocytes compared to controls after a 24-h pulse. Arrowhead (**k**) marks the leading edge of migrating

BrdU-positive enterocyte. **l, m.** Representative images of *Olfm4* (**l**,  $n = 3$ ) and *Crp4* (**m**,  $n = 6$ ) *in situ* hybridizations from control and HFD-fed mice. **n.** No significant difference in the number of jejunal caspase3<sup>+</sup> cells was detected by immunohistochemistry. Images are representative of three separate experiments ( $n = 3$ ); arrowheads indicate representative caspase3<sup>+</sup> enterocytes. **o.** The HFD chow (Research Diets D12492) provides a higher percentage of kilocalories from fat and conversely a lower percentage of kilocalories from protein and carbohydrates compared to a standard chow diet (Labdiet RMH3000). Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments; \* $P < 0.05$ , \*\*\* $P < 0.01$ , \*\*\*\* $P < 0.001$  (Student's  $t$ -tests). Scale bars, 100  $\mu\text{m}$  (**g**, **k**, **n**), 50  $\mu\text{m}$  (**h**, **i**, **k** (inset), **l**, **m**) and 20  $\mu\text{m}$  (**l**, **m**, insets); two separate fields of jejunum (**d**), and at least 15 crypts (**e**), 15 villi (**f**), 10 villi (**g**), 100 cells (**i**), 25 villi (**j**) and 25 crypt–villus units (**n**) were counted per sample in each independent experiment.

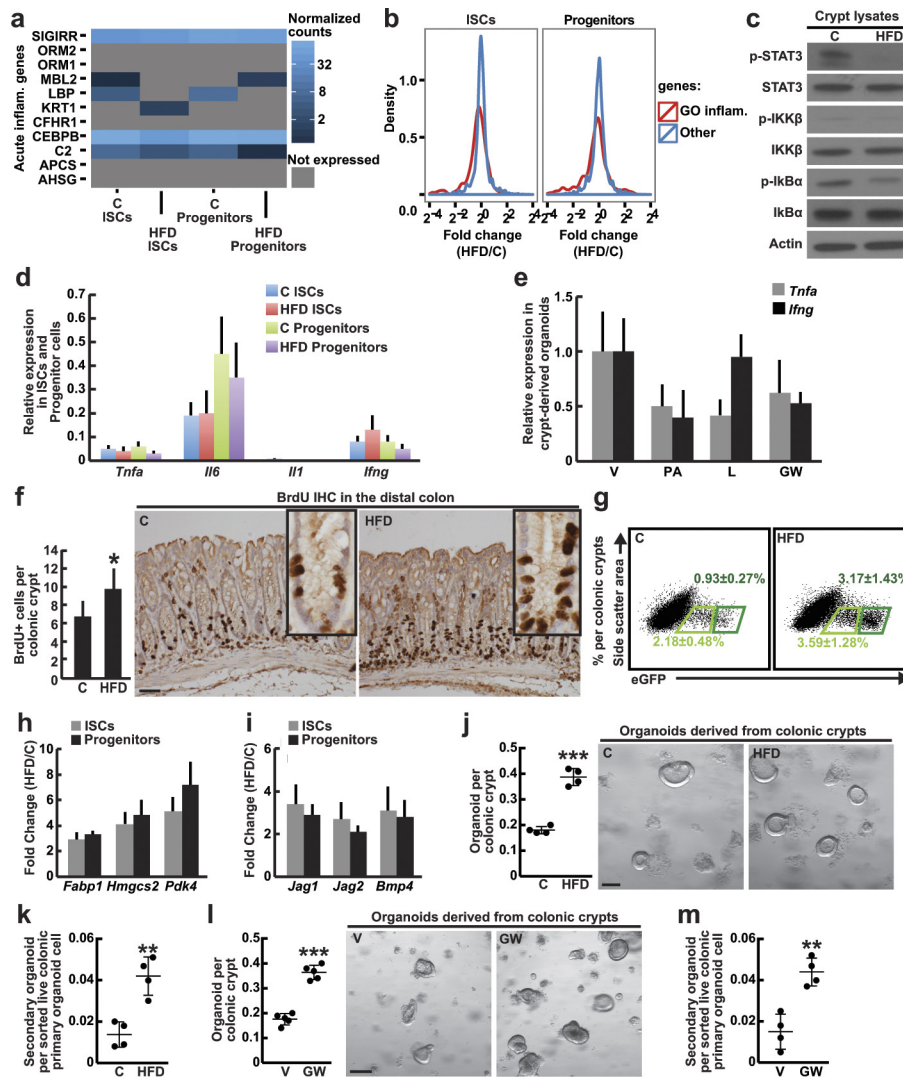




**Extended Data Figure 2 | A HFD and PPAR- $\delta$  signalling have minimal effects on enteroendocrine and goblet cell differentiation but promote intestinal regeneration after 15 Gy irradiation.**

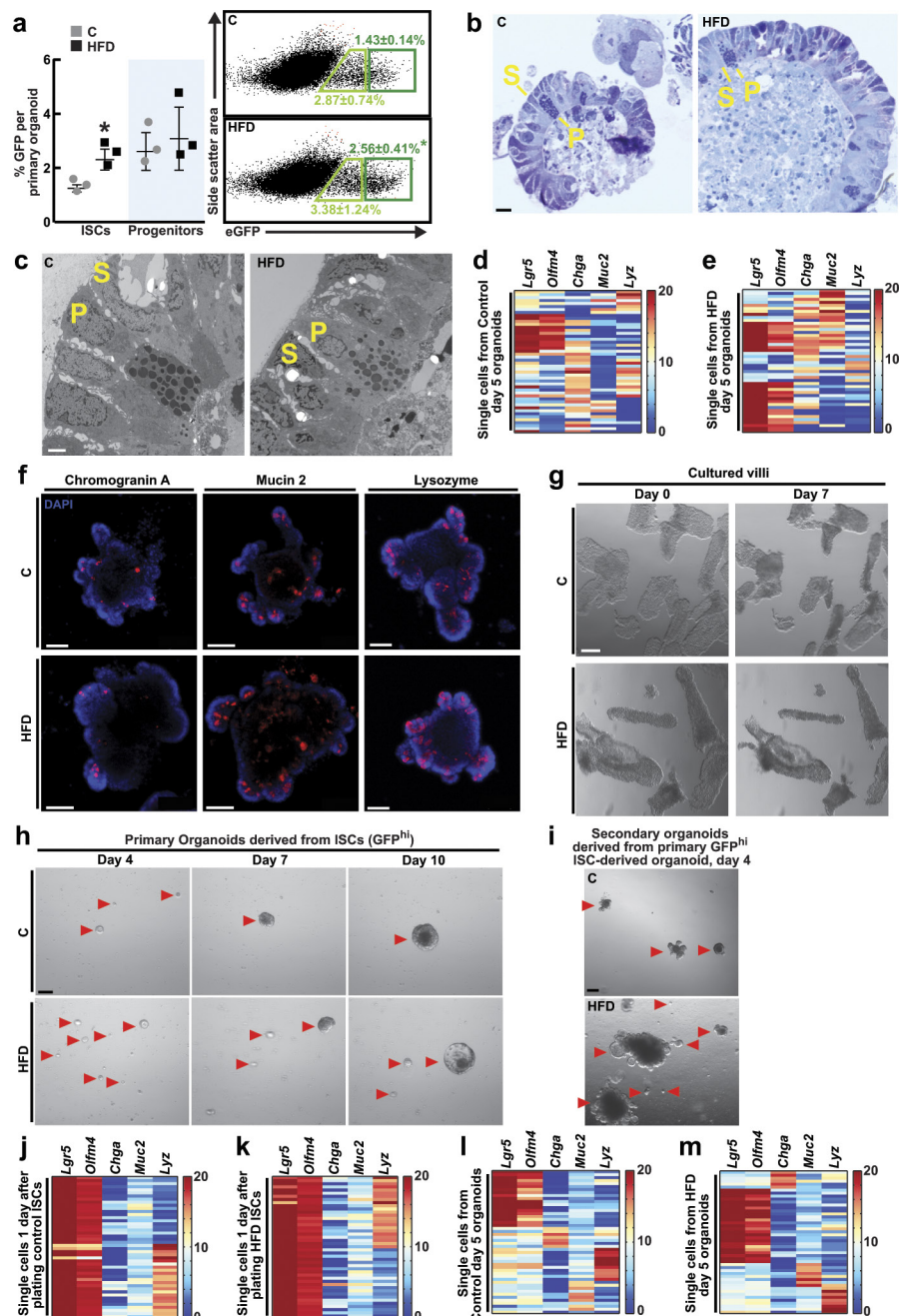
**a, b**, Quantification (**a**,  $n = 3$ ) of immunostains (**b**,  $n = 3$ ) for chromogranin A revealed no difference in the numbers of jejunal enteroendocrine cells (arrowheads) per crypt-villus unit in HFD-fed mice and GW501516-treated mice compared to their respective controls. **c, d**, Quantification (**c**,  $n = 4$ ) of Alcian blue/PAS staining (**d**,  $n = 4$ ) showed no difference in mucinous goblet cells (arrowhead) in HFD-fed and GW501516-treated mice compared to their respective controls. **e, f**, A HFD increased the number of regenerating crypts as measured by

an increased number of crypts containing at least ten Ki67<sup>+</sup> (a marker of proliferation) cells (**e**,  $n = 3$ ) or at least one *Olfm4*<sup>+</sup> cell (**f**,  $n = 3$ ) per 5 mm of jejunum by immunohistochemistry (IHC) or *in situ* hybridization (ISH). Arrows indicate *Olfm4*<sup>+</sup> crypts. **g**, Surviving crypt numbers after ionizing irradiation-induced (XRT) damage. Arrows denote regenerating crypts; asterisks denote aborted crypts ( $n = 3$ ). Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments. Scale bars, 100  $\mu$ m (**b, d**), 50  $\mu$ m (**e–g**) and 20  $\mu$ m (**e, f**, insets); 50 crypt-villus units per sample were analysed (**a, c**) and approximately 50 crypts (**e–g**) were counted per sample in each independent experiment.



**Extended Data Figure 3 | A HFD and fatty acids do not activate inflammatory pathways in intestinal crypts and organoids, while HFD and enforced PPAR- $\delta$  signalling enhance colonic stem-cell function.** **a**, A HFD did not alter the normalized expression levels of inflammatory genes from the GSEA Molecular Signature Database (MSigDB; signature M6557) data set in ISCs and progenitors. **b**, A HFD did not induce differential expression of 'inflammatory response' genes from Gene Ontology (GO; 0006954) in ISCs (Lgr5-GFP<sup>hi</sup>) or progenitors (Lgr5-GFP<sup>low</sup>) compared to control. Fold changes of GO inflammatory response genes are indicated in red, and fold changes for all other genes are indicated in blue. **c**, HFD did not activate the NF- $\kappa$ B or the STAT-3 pathways in the intestinal crypt. Total and phosphorylated protein levels in crypt lysates were assessed by immunoblots ( $n = 3$ ). For western blot source data, see Supplementary Fig. 1. **d**, A HFD did not induce pro-inflammatory gene expression in ISCs (Lgr5-GFP<sup>hi</sup>) or progenitors (Lgr5-GFP<sup>low</sup>). Relative expression levels compared to *Actb* were measured by qRT-PCR ( $n = 5$ ). **e**, *Ex vivo* palmitic acid, lipid mixture or GW501516 treatment did not induce inflammatory gene expression in

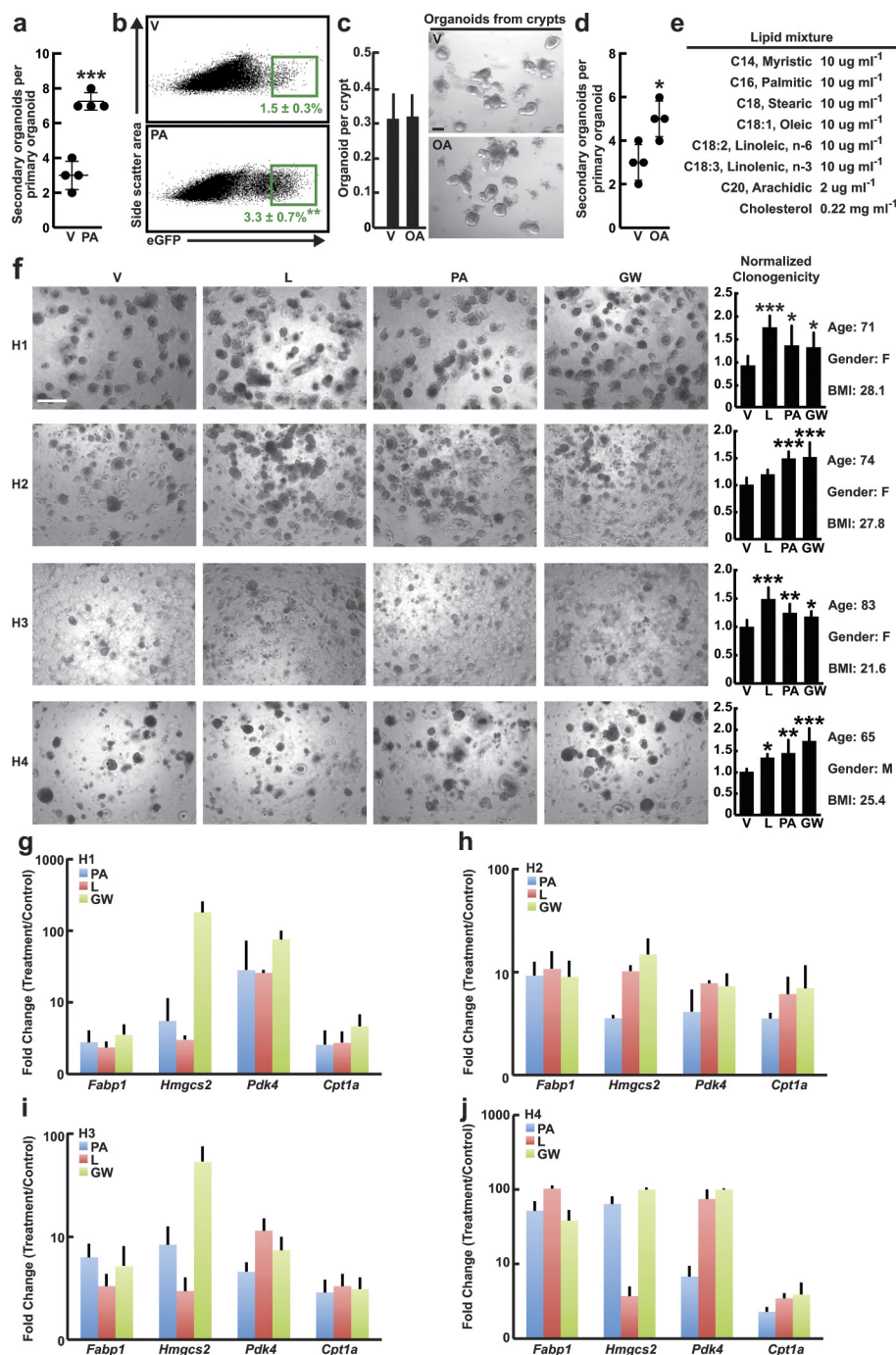
crypt-derived organoids compared to vehicle. Relative expression levels compared to *Actb* were assessed by qRT-PCR ( $n = 4$ , 12 wells per sample were analysed). **f**, A HFD boosted the number of BrdU-labelled cells as measured in distal colonic crypts compared to control (control  $n = 6$ , HFD  $n = 5$ ) after a 4-h pulse. **g**, A HFD increased the frequency of colonic ISCs (Lgr5-GFP<sup>hi</sup>, dark green) and progenitor cells (Lgr5-GFP<sup>low</sup>, light green) ( $n = 8$ ). **h**, **i**, A HFD enhanced PPAR- $\delta$  (**h**) and  $\beta$ -catenin (**i**) target gene expression in colonic ISCs and progenitors. Relative expression levels compared to *Actb* were determined by qRT-PCR ( $n = 5$ , all fold changes are significant with  $P < 0.05$ ). **j**–**m**, Colonic crypts derived from HFD-fed (**j**,  $n = 4$ ; **k**,  $n = 4$ ) and GW501516-treated (**l**,  $n = 5$ ; **m**,  $n = 4$ ) mice demonstrated greater primary and secondary organoid-forming capacity compared to their respective controls. Representative images: day-4 organoids. Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (Student's  $t$ -tests). Scale bars, 50  $\mu$ m (**f**), 100  $\mu$ m (**j**) and 200  $\mu$ m (**l**); 50 crypts per sample were analysed (**f**) in each independent experiment.



**Extended Data Figure 4 | Characterization of HFD crypt and ISC-derived organoids.** **a**, HFD organoids contained higher frequencies of ISCs (*Lgr5-GFP<sup>hi</sup>*) compared to control ( $n=3$ ). **b, c**, Control and HFD organoids demonstrated no differences in morphologic ultrastructure as seen in 1- $\mu$ m sections of control (left) and HFD (right) organoids counterstained with Toluidine Blue (**b**), and electron microscopy images of representative control (left) and HFD (right) organoids (**c**) ( $n=3$ ). **d, e**, Composition of organoids derived from control (**d**) and HFD (**e**) crypts as assessed by single-cell gene expression analysis. Organoids on day 5 contained ISCs (*Lgr5* and *Olfm4*), Paneth cells (*Lyz*), enteroendocrine cells (*Chga*), and goblet cells (*Muc2*). Forty-eight live cells per group were sorted and single-cell gene expression analysis was performed after pre-amplification using corresponding stem-cell and lineage primers (see Methods). **f**, Crypt-derived organoids from control or HFD-fed mice included chromogranin A-, mucin 2- and lysozyme-positive cells as assessed by immunofluorescence (blue = DAPI, red = cell-specific antibody). Images represent two experiments ( $n=2$ ).

**g**, Cultured villi from control and HFD-fed mice lack the ability to form organoids. Images represent two experiments with 6 wells per sample ( $n=2$ ). **h**, ISCs from HFD-fed mice contained greater organoid-forming potential compared to controls. Arrowheads indicate representative organoids at days 4, 7 and 10 of culture ( $n=4$ ). **i**, Individually dissociated HFD primary organoids that were derived from single ISCs possessed more secondary organoid-forming ability than those from controls. ( $n=4$ ). Representative images: day-4 secondary organoids. **j, k**, Single-cell gene expression analysis revealed that ISCs from both control (**j**) and HFD (**k**) mice can beget Paneth cells (*Lyz*) within 24 h in culture (48 cells per group, see Methods). **l, m**, Composition of organoids derived from control (**l**) and HFD (**m**) ISCs (*Lgr5-GFP<sup>hi</sup>*) as assessed by single-cell gene expression analysis (48 cells per group, see Methods). Organoids on day 5 contained ISCs (*Lgr5* and *Olfm4*), Paneth cells (*Lyz*), endocrine cells (*Chga*) and goblet cells (*Muc2*). Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments; \* $P < 0.05$  (Student's  $t$ -tests). Scale bars, 20  $\mu$ m (**b**), 2  $\mu$ m (**c**), 50  $\mu$ m (**f**), 200  $\mu$ m (**g, i**) and 100  $\mu$ m (**h**).

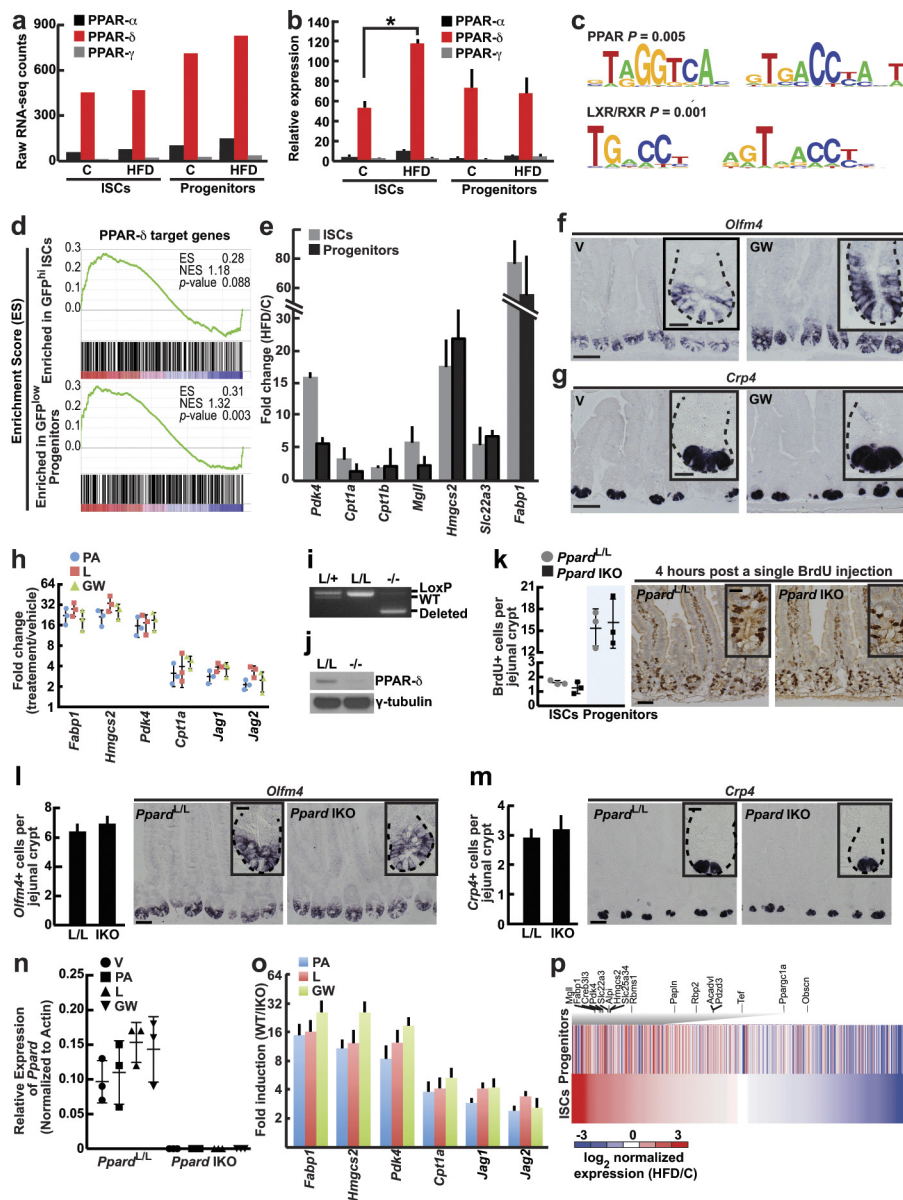




### Extended Data Figure 5 | *Ex vivo* exposure of mouse and human organoids to fatty acids recapitulates aspects of a HFD.

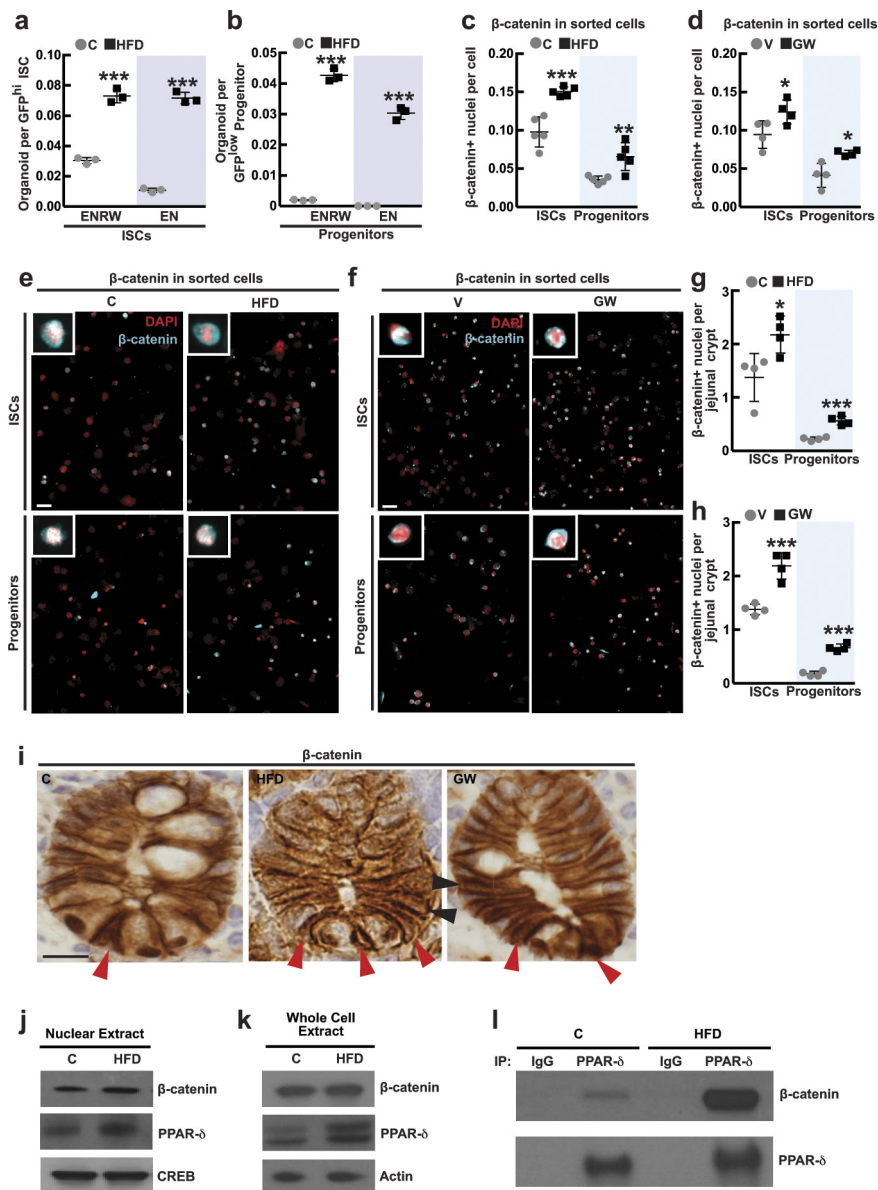
**a, b**, Individually dissociated primary organoids possessed more secondary organoid-forming activity (**a**,  $n = 4$ , the mean number of secondary organoid-forming activity), and contained a higher frequency of *Lgr5*-GFP<sup>hi</sup> ISCs (**b**,  $n = 3$ ) after 4 weeks of treatment with 30  $\mu$ M palmitic acid compared to vehicle. **c**, Exposure of naive crypts to 30  $\mu$ M oleic acid had no effect on primary organoid formation measured at day 7 ( $n = 6$ ). Representative images: day-7 organoids. **d**, Individually dissociated primary organoids possessed more secondary-organoid-forming capacity after 4 weeks of treatment with 30  $\mu$ M oleic acid ( $n = 4$ , the mean number of secondary organoids subcloned from each of 5 primary organoids in 4 independent experiments) compared to vehicle (same vehicle cohort used in **a** and **d**). **e**, Lipid mixture composition (Sigma L0288) as described by the manufacturer. **f**, *Ex vivo* treatment of human-derived small intestinal

crypts (H1–H4) passaged in the presence of lipid mixture, palmitic acid or GW501516 augmented relative clonogenicity compared to vehicle, as shown in representative images from 4 independent experiments. H1:  $n = 10$  (vehicle, palmitic acid, GW501516) and  $n = 6$  (lipid mixture) wells were analysed. H2:  $n = 16$  (vehicle),  $n = 6$  (lipid),  $n = 12$  (palmitic acid) and  $n = 14$  (GW501516) wells were analysed. H3:  $n = 10$  (vehicle),  $n = 12$  (lipid, palmitic acid) and  $n = 8$  (GW501516) wells were analysed. H4:  $n = 7$  (vehicle, GW501516),  $n = 6$  (lipid) and  $n = 9$  (palmitic acid) wells were analysed. Age, gender and BMI are specified. **g–j**, Human crypt-derived organoids after *ex vivo* treatment with palmitic acid, lipid or GW501516 induced PPAR- $\delta$  target gene expression as assessed in passaged cultures with qRT-PCR ( $n = 4$ , 12 wells per sample were analysed, all fold changes are significant,  $P < 0.05$ ). Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (Student's *t*-tests). Scale bars, 100  $\mu$ m (**c**) and 500  $\mu$ m (**f**).



**Extended Data Figure 6 | PPAR- $\delta$  is the predominant PPAR family member expressed in intestinal progenitors and mediates the effects of HFD.** **a**, PPAR- $\delta$  is the most abundant PPAR family member in ISCs (Lgr5-GFP<sup>hi</sup>) and progenitors (Lgr5-GFP<sup>low</sup>) based on RNA-seq data. **b**, Confirmation of PPAR family member mRNA expression levels in ISCs (Lgr5-GFP<sup>hi</sup>) and progenitors (Lgr5-GFP<sup>low</sup>) by qRT-PCR ( $n = 5$ ). **c**, Genes upregulated in HFD ISCs (Lgr5-GFP<sup>hi</sup>) versus control ISCs were enriched in PPAR and LXR/RXR motifs. **d**, GSEA of RNA-seq data identified enrichment of PPAR- $\delta$  targets in ISCs (Lgr5-GFP<sup>hi</sup>) and progenitors (Lgr5-GFP<sup>low</sup>) with a HFD. **e**, Confirmation of induced PPAR- $\delta$  target gene expression in flow-sorted ISCs (Lgr5-GFP<sup>hi</sup>) and progenitors (Lgr5-GFP<sup>low</sup>) by qRT-PCR ( $n = 5$ ). All fold changes were significant,  $P < 0.05$ . **f, g**, Representative images of *Olfm4*<sup>+</sup> ISCs (f) and *Crp4*<sup>+</sup> Paneth cells (g) *in situ* hybridization from vehicle and GW501516-treated mice (f,  $n = 3$ ; g,  $n = 4$ ). **h**, Ex vivo exposure of organoids to palmitic acid, lipid mixture or GW501516 stimulated PPAR- $\delta$  and  $\beta$ -catenin target gene expression ( $n = 3$ , all fold changes were significant,  $P < 0.05$ ). **i, j**, Injection with tamoxifen (4 injections on alternating days) in *Ppard*<sup>L/L</sup>; Villin-CreERT2 mice led to efficient intestinal deletion (IKO) of

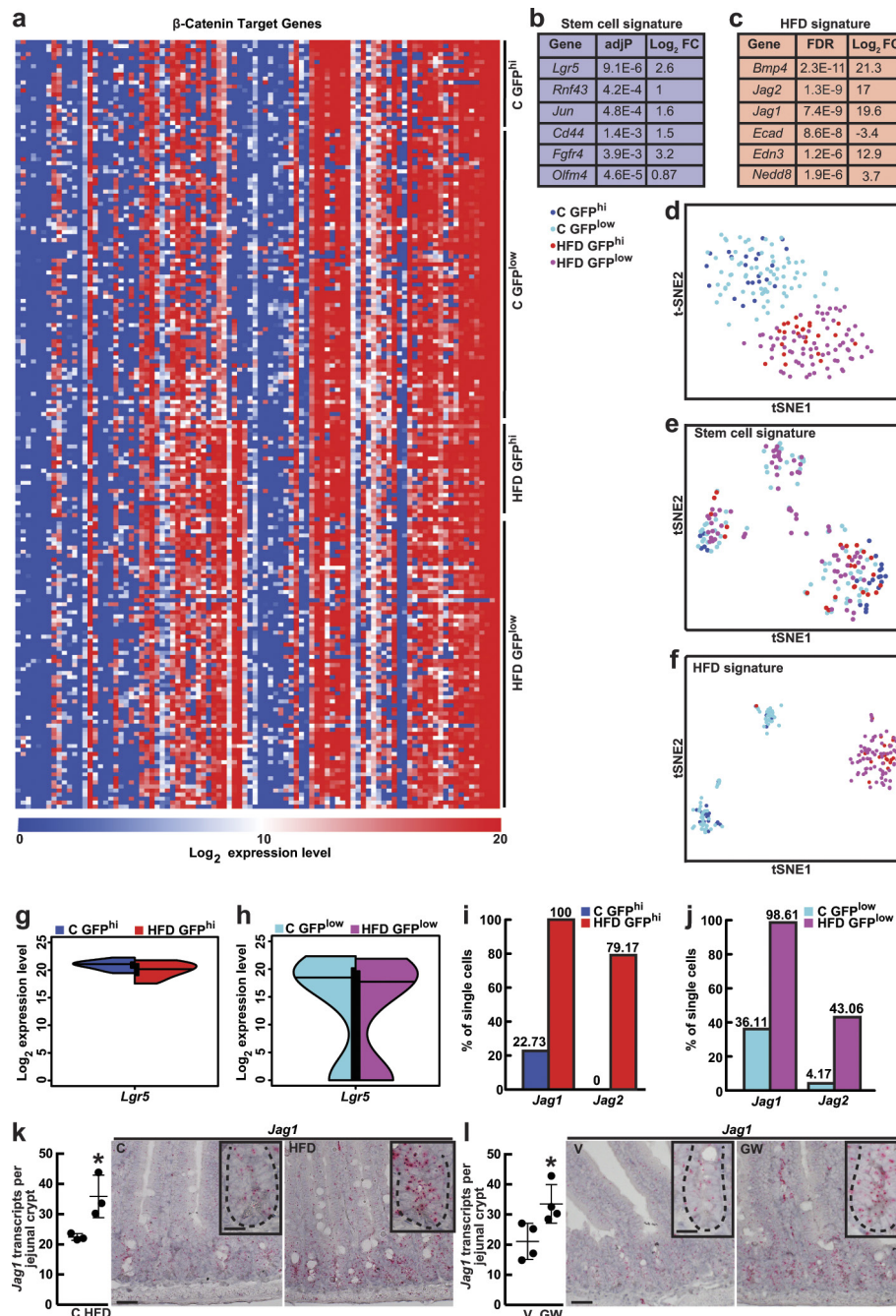
*Ppard* (7 days after the last tamoxifen dose), as assessed by allele-specific deletion PCR (i,  $n = 3$ ) and immunoblot analysis (j,  $n = 3$ ) of crypts. For western blot source data, see Supplementary Fig. 1. **k**, Acute disruption of *Ppard* (8 days after the last tamoxifen dose) did not perturb ISC and progenitor proliferation, as determined 4-h after BrdU administration ( $n = 3$ ). **l, m**, Acute *Ppard* deletion (8 days after the last tamoxifen dose) did not significantly alter *Olfm4*<sup>+</sup> ISCs numbers (L/L:  $n = 5$ , IKO:  $n = 4$ ) (**l**) or *Crp4*<sup>+</sup> Paneth cell ( $n = 5$ ) (**m**) numbers, as assessed by *in situ* hybridization. **n**, Loss of *Ppard* transcripts in *Ppard* IKO organoids was confirmed by qRT-PCR using deletion-specific primers ( $n = 3$ ). **o**, PPAR- $\delta$  is required for the induction of PPAR- $\delta$  and  $\beta$ -catenin target gene expression in secondary organoids after ex vivo palmitic acid, lipid or GW501516 treatment ( $n = 5$ , all fold changes are significant,  $P < 0.05$ ). **p**, Heat map of differentially expressed genes illustrated induction of a PPAR- $\delta$  program in HFD-derived ISCs and progenitors relative to controls. Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments; \* $P < 0.05$  (Student's  $t$ -tests). Scale bars, 50  $\mu$ m (f, g, k-m) and 20  $\mu$ m (insets); 50 crypts per sample were analysed in each independent experiment (f, g, k-m).



**Extended Data Figure 7 | HFD and PPAR- $\delta$  signalling boost nuclear  $\beta$ -catenin localization and activity in intestinal progenitors.** **a, b**, HFD-derived ISCs (**a**, Lgr5-GFP<sup>hi</sup>) and progenitors (**b**, Lgr5-GFP<sup>low</sup>) required less Wnt3a and R-spondin to initiate organoids than control ISCs, as measured by comparing organoid-formation in complete ENRW media, which includes EGF, Noggin, R-spondin and Wnt3a, versus EN media, which includes EGF and Noggin but lacks Wnt3a and R-spondin ( $n = 3$ ). Control-derived progenitors, in contrast to HFD-derived progenitors, rarely formed organoids in either ENRW or EN media. **c–f**, HFD increased nuclear  $\beta$ -catenin localization in flow-sorted ISCs and progenitors from HFD (**c**,  $n = 5$ ) and GW501516-treated (**d**,  $n = 4$ ) mice as determined by immunofluorescence (red, DAPI; cyan, non-phosphorylated  $\beta$ -catenin, CST 8814S). At least 100 cells per sample were quantified. Representative

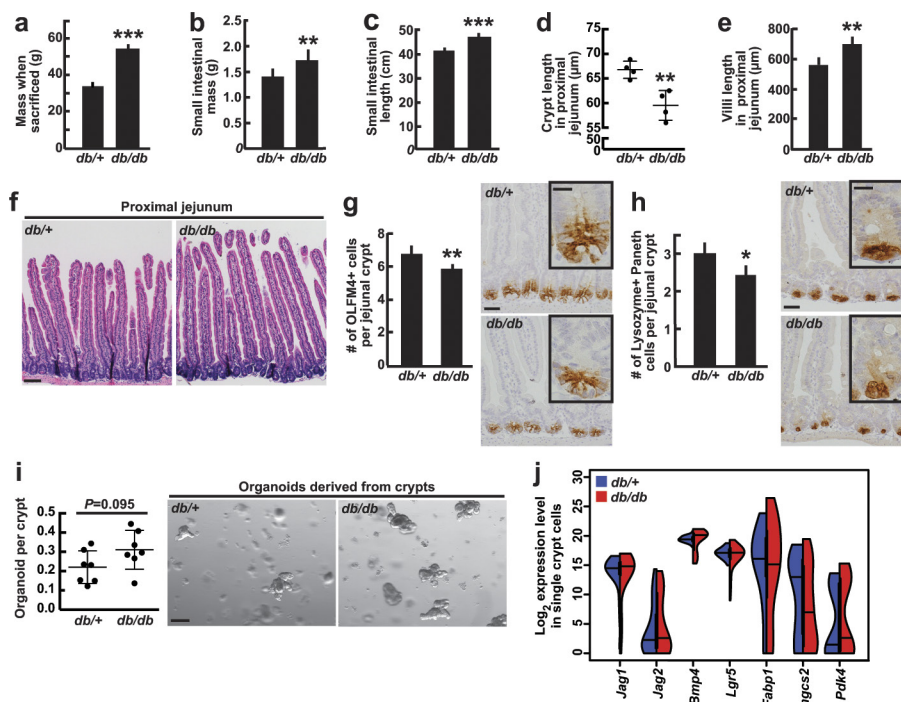
images are shown in **e** and **f**. **g–i**, HFD (**g**) and GW501516 (**h**) treatment increased the numbers of ISCs and progenitors with  $\beta$ -catenin<sup>+</sup> nuclei, as assessed by immunostaining ( $n = 4$  each). Representative images are shown in **i**; arrowheads indicate representative nuclear  $\beta$ -catenin in ISCs (red) and progenitors (black). **j–l**, Association of PPAR- $\delta$  and  $\beta$ -catenin in control and HFD-derived intestinal crypts as shown by immunoprecipitation (IP) ( $n = 3$ ). For western blot source data, see Supplementary Fig. 1. Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (Student's  $t$ -tests). Scale bars, 50  $\mu$ m (**e, f**) and 20  $\mu$ m (**i**); organoid assays: 2–4 wells per sample analysed (**a, b**), 50 crypts per sample were analysed in each independent experiment (**g, h**).





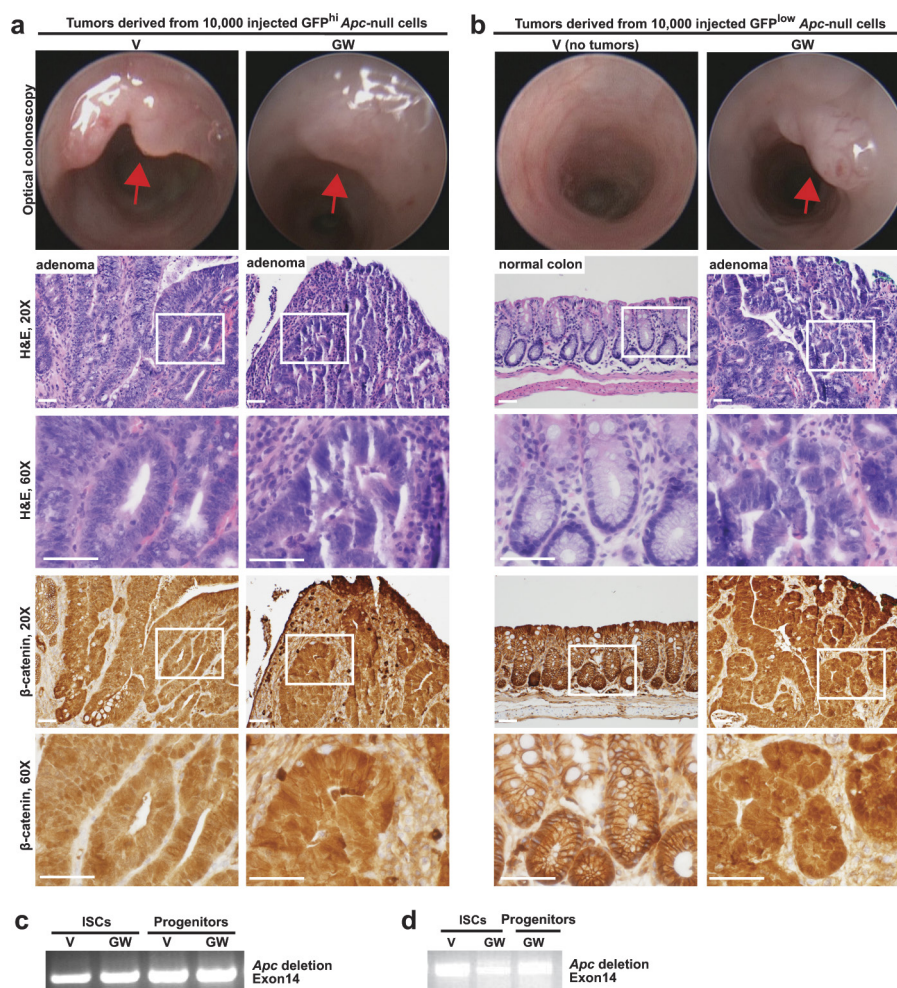
**Extended Data Figure 8 | HFD-mediated alterations in  $\beta$ -catenin target gene expression in single ISCs and progenitors.** **a**, Heat map representation of  $\beta$ -catenin target gene expression in single ISCs (Lgr5-GFP<sup>hi</sup>, 24 cells) and progenitors (Lgr5-GFP<sup>low</sup>, 72 cells) (see Methods). **b**, Stem-cell signature genes were identified by comparing target gene expression in control ISCs (Lgr5-GFP<sup>hi</sup>) to control progenitors (Lgr5-GFP<sup>low</sup>). **c**, HFD signature genes were identified by comparing target gene expression in HFD ISCs to control ISCs (Lgr5-GFP<sup>hi</sup>). **d**, t-Distributed stochastic neighbour embedding (tSNE) analysis of single cells using all  $\beta$ -catenin target genes. **e**, tSNE analysis of single cells using stem-cell signature genes. **f**, tSNE analysis of single cells using HFD signature genes. **g**, **h**,  $Lgr5$  expression was similar in HFD ISCs

(Lgr5-GFP<sup>hi</sup>) (**g**) and progenitors (Lgr5-GFP<sup>low</sup>) (**h**) as compared to their respective controls. **i**, **j**, HFD increased the percentage of ISCs (Lgr5-GFP<sup>hi</sup>) (**i**) and progenitors (Lgr5-GFP<sup>low</sup>) (**j**), with increased  $Jag1$  and  $Jag2$  expression compared to their respective controls. **k**, **l**, HFD (**k**,  $n = 3$ ) and GW501516 treatment (**l**,  $n = 4$ ) augmented  $Jag1$  expression compared to control and vehicle treatments, respectively, as assayed by single-molecule *in situ* hybridization:  $Jag1$  is broadly expressed throughout the crypt. Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments; \* $P < 0.05$  (Student's  $t$ -tests). Scale bars, 50  $\mu$ m (**k**, **l**) and 20  $\mu$ m (insets); more than 50 crypts per sample were analysed in each independent experiment (**k**, **l**). See Supplementary Information for raw gene expression data.



**Extended Data Figure 9 | Characterization of obese *db/db* mouse intestines.** **a–f**, At 4–5 months of age, homozygous *db/db* mice gained on average 50% more mass (**a**,  $n=9$ ), had increased small intestinal mass and length (**b**, **c**,  $n=9$ ), shallower crypts (**d**,  $n=4$ ) and longer villi (**e**, **f**,  $n=5$ ) than control *db/+* mice. **g**, **h**, Immunostains for OLFM4 ( $n=6$ ) and lysozyme ( $n=6$ ) revealed a slight reduction in the number of Olfm4<sup>+</sup> ISCs and Paneth cells, respectively, in *db/db* mice compared to *db/+* controls. **i**, Organoid-forming capacity of *db/db* crypts was higher ( $P=0.095$ ) than *db/+* controls ( $n=7$ ). **j**, Single-cell gene expression analysis revealed no induction of PPAR- $\delta$  or  $\beta$ -catenin target gene

expression in live, enriched stem and progenitor cells that are depleted of secretory cells (7-AAD<sup>−</sup>Epcam<sup>+</sup>CD24<sup>−</sup>c-kit<sup>−</sup> cells, 48 cells per group; see Methods) from *db/db* intestines compared to control intestines. Unless otherwise indicated, data are mean  $\pm$  s.d. from  $n$  independent experiments; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (Student's  $t$ -tests). Scale bars, 100  $\mu\text{m}$  (**f**), 50  $\mu\text{m}$  (**g**, **h**), 20  $\mu\text{m}$  (**g**, **h**, insets) and 200  $\mu\text{m}$  (**i**); and at least 30 crypts (**d**), 20 villi (**e**) and 100 crypts (**g**, **h**) were assessed per sample in each independent experiment. All *db/db* and *db/+* mice were fed a standard chow diet.



**Extended Data Figure 10 | PPAR- $\delta$  activation bestows adenoma-initiating capacity to *Apc*-null progenitors.** **a, b,** Representative optical endoscopy images (top) from Fig. 5, with H&E (middle) and  $\beta$ -catenin (immunohistochemistry, bottom) sections of adenomas derived from orthotopic transplantation of *Apc*-null ISC (a, Lgr5-GFP<sup>hi</sup>) and progenitors (b, Lgr5-GFP<sup>low</sup>) from vehicle- and GW501516-treated mice 4 days after *Apc* deletion. Tumours exhibited hyperchromasia, lack of maturation, nuclear crowding and nuclear  $\beta$ -catenin positivity. Two

independent pathologists blinded to treatment groups interpreted the results. **c, d**, *Apc* deletion was confirmed in sorted small intestinal ISCs and progenitors from vehicle- and GW501516-treated *Apc<sup>L/L</sup>; Lgr5-EGFP-IRES-CreERT2* mice 4 days after tamoxifen administration (**c**,  $n = 3$ ) and in isolated tumours (**d**,  $n = 3$ ) by PCR amplification, using allele-specific deletion primers targeting exon 14. Unless otherwise indicated,  $n$  represents independent experiments. Scale bars (**a, b**), 50  $\mu\text{m}$  (20 $\times$ ) and 20  $\mu\text{m}$  (60 $\times$ ).



# Priming and polymerization of a bacterial contractile tail structure

Abdelrahim Zoued<sup>1\*</sup>, Eric Durand<sup>1,2,3,4,5\*</sup>, Yannick R. Brunet<sup>1†</sup>, Silvia Spinelli<sup>2,3</sup>, Badreddine Douzi<sup>1,2,3</sup>, Mathilde Guzzo<sup>6</sup>, Nicolas Flaugnatti<sup>1</sup>, Pierre Legrand<sup>7</sup>, Laure Journet<sup>1</sup>, Rémi Fronzes<sup>4,5</sup>, Tãm Mignot<sup>6</sup>, Christian Cambillau<sup>2,3</sup> & Eric Cascales<sup>1</sup>

**Contractile tails are composed of an inner tube wrapped by an outer sheath assembled in an extended, metastable conformation that stores mechanical energy necessary for its contraction. Contraction is used to propel the rigid inner tube towards target cells for DNA or toxin delivery. Although recent studies have revealed the structure of the contractile sheath of the type VI secretion system, the mechanisms by which its polymerization is controlled and coordinated with the assembly of the inner tube remain unknown. Here we show that the starfish-like TssA dodecameric complex interacts with tube and sheath components. Fluorescence microscopy experiments in enteroaggregative *Escherichia coli* reveal that TssA binds first to the type VI secretion system membrane core complex and then initiates tail polymerization. TssA remains at the tip of the growing structure and incorporates new tube and sheath blocks. On the basis of these results, we propose that TssA primes and coordinates tail tube and sheath biogenesis.**

Contractile injection machines are nano-structures evolved to deliver macromolecules into target cells<sup>1</sup>. These machines have been elaborated for different purposes such as the injection of DNA into host cells in the case of bacteriophages, for the delivery of protein effectors into bacterial or eukaryotic cells in the case of R-pyocins, *Photorhabdus* virulence cassettes, anti-feeding prophages or type VI secretion systems (T6SS) or for inducing metamorphosis in invertebrates<sup>1–6</sup>. These machines include a tubular edifice called a tail<sup>1,7,8</sup>. The tail is essentially composed of a rigid inner tube wrapped by a contractile structure—the sheath—that is assembled in an extended conformation that stores mechanical energy necessary for its contraction and to propel the inner tube towards the target<sup>5</sup>. The tail is assembled on the baseplate that varies in terms of composition and number of subunits; however, a minimal baseplate consists of the hub protein surrounded by wedges<sup>1,7,8</sup>. The baseplate is not only the platform for the assembly of the tube/sheath, but also an important component of the signalling cascade that triggers sheath contraction<sup>1,8</sup>. Tails are usually completed by terminator proteins that stabilize the sheath and maintain tube and sheath together at the distal end to prevent energy dissipation during sheath contraction and to permit proper ejection of the inner tube<sup>8–10</sup>.

The T6SS is composed of a contractile structure anchored to the cell envelope by the TssJLM membrane complex that serves as a docking station as well as a channel for the passage of the inner tube during sheath contraction<sup>11–14</sup> (Extended Data Fig. 1a). The contractile structure is composed of the tail tube made up of stacks of Hcp hexameric rings, wrapped by a sheath-like structure consisting of the TssB and TssC subunits (Extended Data Fig. 1a)<sup>14</sup>. During T6SS biogenesis, the assembly of the tube and sheath are coordinated: the insertion of a tube ring immediately preceding that of a sheath block<sup>15</sup>. This tail polymerizes on a baseplate-like complex composed of the VgrG hub and the TssE, TssF, TssG and TssK subunits<sup>16–19</sup> (Extended Data Fig. 1a). The TssBC sheath polymerizes in tens of seconds to build an ~600-nm long structure that

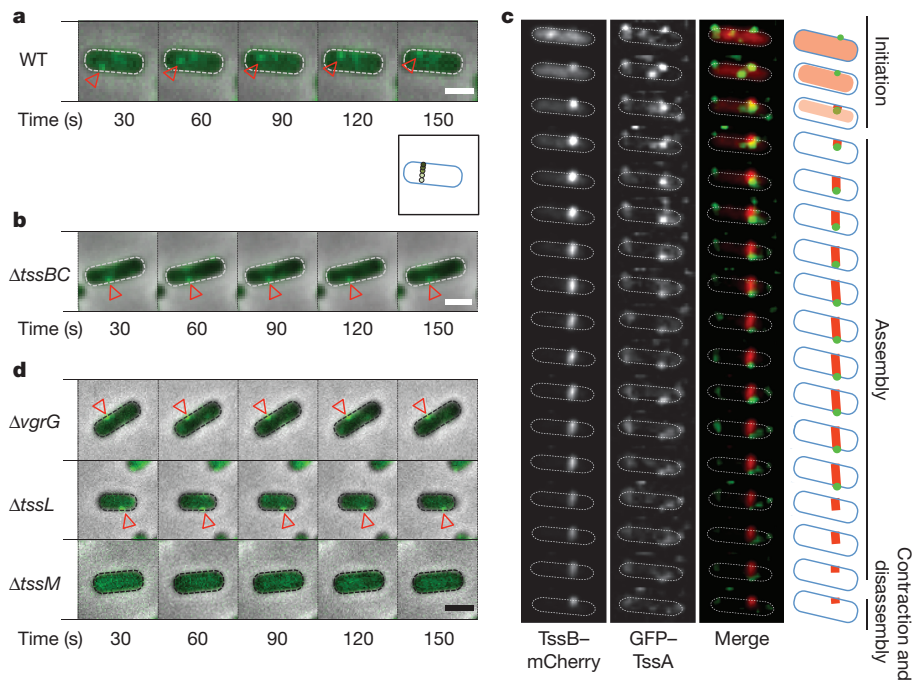
contracts in a few milliseconds<sup>17</sup>. Contraction of the sheath propels the Hcp inner tube towards the target cell, like a ‘nano-crossbow’<sup>14</sup>, and is responsible for the delivery of toxin effectors, as it correlates with lysis of the competitor bacterium<sup>20,21</sup>. Recent cryo-electron microscopy studies have revealed the atomic structure of the T6SS sheath in its contracted conformation<sup>22,23</sup>. The sheath is a helical structure composed of 6-TssB/TssC heterodimer strands, each heterodimer being stabilized by an intra- and inter-strand handshake domain<sup>23</sup>. In addition, a cryo-electron microscopy study of the pyocin R2 has provided information regarding the atomic structure of this contractile nanotube in its extended conformation and on how it interacts with the inner tube<sup>2</sup>. Although the general mechanism of T6SS assembly and the structure of the T6SS sheath are now well documented, critical details are missing, such as how the polymerization of the sheath is controlled, how tube and sheath assembly is coordinated and how tail polymerization is stopped.

## TssA initiates tail tube/sheath polymerization

During T6SS tail biogenesis, the recruitment and assembly of Hcp hexamers and TssBC sheath blocks should be coordinated and the tail tube and sheath should be firmly attached together at the distal end to allow proper tube throwing during contraction. We therefore hypothesized that at least one of the T6SS core proteins must be required to coordinate and/or terminate Hcp/TssBC tail assembly. Such candidate subunit(s) should interact with both the tube protein (Hcp) and with at least one component of the sheath (TssB and/or TssC). We therefore performed a systematic bacterial two-hybrid analysis in which Hcp, TssB and TssC were used as baits to identify prey partners within T6SS subunits. Extended Data Fig. 1b shows that a number of baseplate components (TssE, TssF, TssG and VgrG) interact with either Hcp or TssC. However, a unique protein, TssA (GenBank accession number: 284924261), interacts with both tube and sheath components. Recent

<sup>1</sup>Laboratoire d'Ingénierie des Systèmes Macromoléculaires, Institut de Microbiologie de la Méditerranée, CNRS UMR7255, Aix-Marseille Université, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France. <sup>2</sup>Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique, UMR 7257, Campus de Luminy, Case 932, 13288 Marseille Cedex 09, France. <sup>3</sup>Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille Université, UMR 7257, Campus de Luminy, Case 932, 13288 Marseille Cedex 09, France. <sup>4</sup>G5 Biologie structurale de la sécrétion bactérienne, Institut Pasteur, 25–28 rue du Docteur Roux, 75015 Paris, France. <sup>5</sup>UMR 3528, CNRS, Institut Pasteur, 25–28 rue du Docteur Roux, 75015 Paris, France. <sup>6</sup>Laboratoire de Chimie Bactérienne, Institut de Microbiologie de la Méditerranée, CNRS UMR7283, Aix-Marseille Université, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France. <sup>7</sup>Synchrotron Soleil, L'Orme des merisiers, Saint-Aubin BP48, 91192 Gif-sur-Yvette Cedex, France. <sup>†</sup>Present address: Department of Microbiology and Immunobiology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

\*These authors contributed equally to this work.



**Figure 1 | *In vivo* imaging of sfGFP-TssA.** **a**, TssA localizes in mobile foci. Fluorescence microscopy time-lapse recording of wild-type (WT) EAEC cells producing sfGFP-TssA. Individual images were taken every 30 s. The localization of TssA is indicated by the red arrowhead. A schematic diagram representing the localization of TssA (from light to dark green as a function of time) is shown in the inset. Mean square displacement and kymograph analyses are shown in Extended Data Fig. 3a, b. Scale bar, 1  $\mu$ m. **b**, TssA forms static foci in absence of the TssBC sheath proteins. Fluorescence microscopy time-lapse recording of  $\Delta tssBC$  cells producing sfGFP-TssA. **c**, TssA is associated with the distal end of the TssBC sheath during elongation. Fluorescence microscopy time-lapse recording of wild-type EAEC cells producing sfGFP-TssA and TssB-mCherry. The mCherry

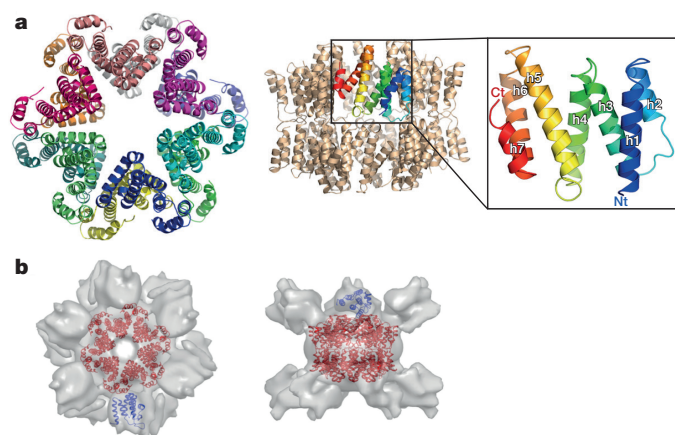
channel (left), GFP channel (middle) and merge channels (right) are shown. Individual images (from top to bottom) were taken every 30 s. The initiation, polymerization and contraction/disassembly stages of the T6SS sheath dynamics are indicated on the right, with a schematic diagram of the observed events. **d**, TssA initial localization requires TssM but not VgrG or TssL. Fluorescence microscopy time-lapse recording of the indicated  $\Delta vgrG$ ,  $\Delta tssL$  or  $\Delta tssM$  cells producing sfGFP-TssA. Individual images were taken every 30 s. Scale bar, 1  $\mu$ m. Recordings of sfGFP-TssA in  $\Delta tssE$ ,  $\Delta tssF$ ,  $\Delta tssG$ ,  $\Delta tssK$  and  $\Delta hcp$  cells are shown in Extended Data Fig. 3e. Large fields, statistical analyses and kymographs analyses are shown in Extended Data Fig. 3f–h, respectively.

data showed that TssA is required for proper formation of the Hcp tube<sup>19</sup>, arguing against a role of TssA as a terminator-only protein. In agreement with this conclusion, fluorescence microscopy experiments show that T6SS sheaths do not assemble in *tssA* cells (Extended Data Fig. 1c). In addition, Hcp tube proteins are not released from *tssA* cells (Extended Data Fig. 1d). Collectively, these data suggest a critical role of TssA as a regulator of T6SS tail biogenesis.

Given that T6SS tube/sheath assembly is initiated on the baseplate complex which is docked at the cytoplasmic side of the TssJLM complex<sup>19</sup>, we hypothesized that TssA interacts with baseplate or membrane complex components. First, fractionation experiments showed that the TssA protein mainly localizes in the cytoplasm but a significant amount of the protein is associated with the membrane fraction (Extended Data Fig. 1e). TssA association with the inner membrane is probably dependent on the T6SS membrane complex as co-purification and gel-filtration experiments showed that TssA binds to the detergent-solubilized TssJLM complex (Extended Data Fig. 2a, b). Negative-stain electron microscopy of the TssJLM-TssA complex further demonstrated the presence of an  $\sim 300$  Å-large complex associated to the cytoplasmic base of the TssJLM rocket-like structure (Extended Data Fig. 2c, d). To gain further information on TssA localization and dynamics, we fused TssA to a fluorescent reporter, the superfolder green fluorescent protein (sfGFP). sfGFP was inserted at the *tssA* locus on the chromosome, to engineer cells producing a functional sfGFP-TssA chimera protein. Time-lapse microscopy recordings showed that TssA does not distribute randomly but rather assembles 1–3 discrete foci, located close to the membrane (Fig. 1a). Most of these foci are not fixed and show directional movement (Extended Data Fig. 3a). Kymographic analyses confirmed that TssA foci move with a unidirectional trajectory, at a constant velocity (see the schematic representation in the inset of

Fig. 1a, and the kymograph in Extended Data Fig. 3b). Based on these trajectories, we wondered whether TssA foci might be pushed by the elongation of the sheath. In  $\Delta tssBC$  cells, TssA foci were still assembled at close proximity to the membrane, but their dynamics were completely abolished: the TssA foci remained static (Fig. 1b and Extended Data Fig. 3a). This experiment defined that the TssBC sheath does not interfere with TssA recruitment and primary localization but rather pushes the TssA cluster during its elongation. We further monitored TssB-mCherry and sfGFP-TssA dynamics. Time-lapse recordings and kymographic analyses confirmed that TssA-containing complexes assemble first close to the membrane and are then pushed by the sheath towards the opposite side of the bacterium (Fig. 1c and Extended Data Fig. 3b). In addition, fluorescence lifetime imaging microscopy (FLIM) assays demonstrated that TssA molecules do not appear to turn over between the foci and the intracellular pool, therefore suggesting that the same TssA or TssA-containing complex remains at the distal end of the sheath (Extended Data Fig. 3c, d). However, because sheath contraction is a very fast event that occurs in a few milliseconds<sup>17</sup> and is immediately followed by ClpV-mediated disassembly<sup>24</sup>, our experiments could not define whether TssA remains associated to the sheath during contraction.

To provide insights onto the assembly of the T6SS and particularly on the early events before tube/sheath elongation, we tested the localization and dynamics of sfGFP-TssA in various mutant backgrounds. The biogenesis of the T6SS begins with the initial positioning of the TssJ outer membrane lipoprotein and progresses with the sequential recruitment of the inner membrane TssM and TssL proteins<sup>13</sup>, and then the TssEFGK-VgrG baseplate<sup>19</sup>. Deletion of baseplate components or Hcp did not affect TssA localization but abolished its dynamics (Fig. 1d and Extended Data Fig. 3e–h). By contrast, the sfGFP-TssA fluorescence



**Figure 2 | High-resolution structures of TssA domains.** **a**, X-ray structure of the C-terminal domain of TssA (TssA<sub>Ct</sub>) (PDB: 4YO5). Top (left panel) and side (right panel) views of the TssA<sub>Ct</sub> dodecamer structure, shown in ribbon representation with each monomer differently coloured. The inset of the right panel shows the rainbow coloured (blue to red from the N terminus) structure of one TssA<sub>Ct</sub> monomer. The consecutive  $\alpha$ -helices are numbered h1 to h7. The crystal structure of TssA<sub>Nt12</sub> is shown in Extended Data Fig. 6f. **b**, Fitting of the TssA<sub>Ct</sub> (red ribbon) and TssA<sub>Nt12</sub> (blue ribbon) X-ray structures into the TssA electron microscopy reconstruction top (top panel) or side (bottom panel) views (EMD-3282; grey volume). Scale bar, 10 nm. SAXS and negative-stain electron microscopy models of TssA are shown in Extended Data Fig. 5h–j and 5p–r, respectively. Scale bar, 5 nm.

was diffuse in *tssM* cells but remained clustered in *tssL* cells (Fig. 1d and Extended Data Fig. 3e–h). Taken together, these results show that TssA is recruited in the early stages of T6SS biogenesis, after formation of the TssJ–TssM complex (Extended Data Fig. 3i). The interaction of TssA with the TssJM complex was further confirmed by co-purification experiments (Extended Data Fig. 2a).

To identify TssA additional partners, the interaction of TssA with all the T6SS soluble core components and the soluble domains of the TssL, TssM and TssJ proteins, was first tested by a bacterial two-hybrid approach. Extended Data Fig. 4a shows that in addition to Hcp and TssC, and the previously described TssA–TssK interaction<sup>25</sup>, TssA interacts with the TssE and VgrG baseplate subunits. To validate these results by an alternative approach, native TssA, Hcp, TssE, VgrG and the TssBC complex were purified and the interactions were assessed by surface plasmon resonance (Extended Data Fig. 4b–e). In the four cases, we observed interactions between the two partners, with dissociation constants ranging from  $\sim 2$  to  $\sim 50 \mu\text{M}$  (Extended Data Fig. 4b–e). The affinities of these complexes are rather low but have to be replaced in the context of multiple interactions that probably act synergistically. The interaction of TssA with Hcp or TssBC have higher affinities compared to that of TssA with VgrG and TssE, whereas TssA dissociates more rapidly from VgrG or TssE than from Hcp or TssBC (Extended Data Fig. 4b–e). This suggests that the TssE and VgrG baseplate proteins may first interact to TssA and that these interactions will be displaced by the tube and sheath components (Hcp, TssBC).

Based on the interaction network and on the fluorescence microscopy recordings, we conclude that positioning of TssA on the membrane complex recruits the baseplate and initiates polymerization of the tube and sheath. However, TssA remains associated with the distal end of the polymerized structure during the elongation process.

### Structural organization of the TssA protein

The electron microscopy analyses of the TssJLM–TssA complex (Extended Data Fig. 2d) suggest that TssA assembles an  $\sim 300 \text{ \AA}$ -large complex. A  $6 \times$  His-tagged thioredoxin–TssA fusion (TRX–TssA) was produced, purified to homogeneity by ion-metal affinity and size-exclusion chromatographies (Extended Data Fig. 5a). Gel-filtration

(Extended Data Fig. 5b) and on-line multi-angle laser light scattering/quasi-elastic light scattering/absorbance/refractive index (MALS/QELS/UV/RI, Extended Data Fig. 5c) analyses defined a mass of 891 kDa, which corresponds to the mass of a TRX–TssA dodecamer (theoretical mass = 888 kDa). Crystallization attempts with the purified full-length TssA protein obtained after tag cleavage by the TEV protease failed. We therefore examined the protein sample using small-angle X-ray scattering (SAXS, Extended Data Fig. 5d–j) and electron microscopy after negative staining (Extended Data Fig. 5k–r). The  $\sim 19 \text{ \AA}$  resolution single-particle reconstruction of TssA from the electron micrographs (Extended Data Fig. 5k–n) showed that TssA assembles two stacked hexamers with arm-like short extensions (Extended Data Fig. 5p–r). However, we noted that the electron microscopy density does not account for the complete mass of a TssA dodecamer and therefore suspected that the arms may represent a flexible domain, shortened by the averaging procedure. This was confirmed by SAXS studies demonstrating that TssA is composed of a central hexameric core bearing six long arms yielding a starfish-like structure (Extended Data Fig. 5h–j and Extended Data Table 1). Whereas the SAXS model allows better visualization of the arm length compared to the electron microscopy reconstruction, its low resolution impairs the visual separation of the dimeric arms.

Limited proteolysis of full-length TssA using proteinase K yielded three stable fragments of  $\sim 45$ ,  $\sim 33$  and  $\sim 19 \text{ kDa}$  (Extended Data Fig. 6a) corresponding to the N-terminal 1–392, C-terminal 221–531 and 393–531 regions of TssA, respectively (Extended Data Fig. 6b). On the basis of this result, two domains corresponding to fragments 1–400 (TssA<sub>Nt</sub>, theoretical molecular weight, 44.2 kDa) and 395–531 (TssA<sub>Ct</sub>, theoretical molecular weight, 15.1 kDa) were designed. Bacterial two-hybrid experiments showed that these two domains oligomerize independently (Extended Data Fig. 6c). MALS/QELS/UV/RI analyses of the purified TssA<sub>Nt</sub> and TssA<sub>Ct</sub> domains revealed that the N-terminal fragment is a dimer in solution (Extended Data Fig. 6d), whereas the C-terminal domain is dodecameric (Extended Data Fig. 6e).

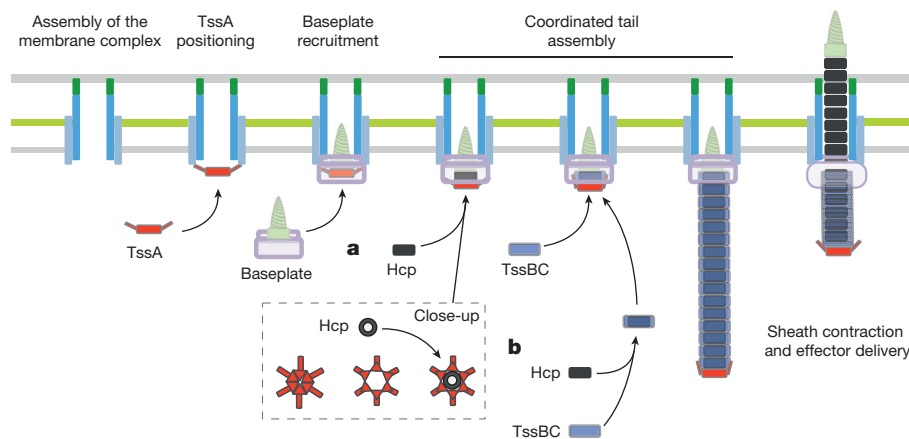
The N-terminal and C-terminal domain crystallized, and 3.37  $\text{\AA}$  and 3.35  $\text{\AA}$  resolution data sets were collected, respectively (Extended Data Table 2). In agreement with the MALS/QELS/UV/RI data, TssA<sub>Nt</sub> is a dimer in the crystal. The peptidic chain starts at residue 221 (instead of 1) and stops at residue 377 (instead of 401). This fragment, named TssA<sub>Nt12</sub>, results from protein cleavage during crystallization. The TssA<sub>Nt12</sub> domain consists of three pairs of helices (1–2, 3–4 and 5–6) structured as a bundle followed by a seventh helix perpendicular to the others (Extended Data Fig. 6f, g). The TssA<sub>Ct</sub> structure features two stacked head-to-head hexamers 30  $\text{\AA}$  thick, with an external diameter of 100  $\text{\AA}$  (Fig. 2a). The overall form of the hexamer is unique and resembles that of a hexaflexagon, a kaleidocycle with six triangular wedges that contact each other via  $\alpha$ -helix hinges (Fig. 2a). Docking of these X-ray structures onto the starfish-shaped electron microscopy and SAXS reconstructions (Fig. 2b and Extended Data Fig. 7a–n) showed that the TssA<sub>Ct</sub> diameter and width coincide exquisitely with the central core of TssA, whereas TssA<sub>Nt12</sub> dimers fit in the arms at close proximity to the TssA<sub>Ct</sub> central core. Large density regions remain available at the extremity of the arms and probably correspond to TssA<sub>Nt1</sub> domains.

To understand the contribution of the TssA central core and arm domains to the T6SS assembly mechanism, we tested interactions between these domains and TssA partners. Bacterial two-hybrid analyses demonstrated that TssA<sub>Nt</sub> interacts with TssBC and TssE whereas TssA<sub>Ct</sub> is sufficient to make contacts with Hcp and VgrG (Extended Data Fig. 6h). SPR analyses further confirmed the TssA<sub>Nt</sub>–TssBC and TssA<sub>Ct</sub>–Hcp interactions with  $K_d$  values of  $\sim 4.2$  and  $\sim 48 \mu\text{M}$ , respectively (Extended Data Fig. 6i, j).

### Closing remarks

In this work we identified TssA as a partner of both T6SS inner tube and sheath. We combined structural, biochemical, functional and microscopy analyses to reveal the specific role of TssA subunit during





**Figure 3 | Model of the assembly of the type VI secretion system.**

**a, b,** Schematic representation of the different stages of the assembly and mechanism of action of the T6SS (from left to right) highlighting the role of TssA. The TssA dodecamer (red) is recruited to the T6SS membrane complex. A negative-stain electron microscopy image of the TssJLM–TssA complex is shown in Extended Data Fig. 2d. TssA then recruits the baseplate and initiates polymerization of the tail by the incorporation of

Hcp (black rectangles) and TssBC (blue rectangles) building blocks (**a**) or the incorporation of Hcp–TssBC building blocks (**b**). During the polymerization, TssA remains at the distal end of the structure. TssA is released after sheath contraction. Docking experiments of TssA at the distal extremity of the extended and contracted sheath are shown in Extended Data Fig. 8a–c. A dynamic representation of this working model is shown in Supplementary Video 1.

T6SS biogenesis. TssA is recruited to an early stage of T6SS biogenesis, positions the baseplate, and initiates and guides tail tube/sheath polymerization.

A TssA dodecamer is recruited to the TssJM complex and clusters in discrete foci that represent the site of assembly of the T6SS tail tube and sheath. Previous studies have shown that TssA is not required for proper recruitment of TssL to the TssJM complex<sup>13</sup>, and hence the T6SS assembly pathway is branched on TssM (Extended Data Fig. 3i). Protein–protein interaction and fluorescence microscopy analyses defined that TssA recruits two components of the baseplate (VgrG and TssE). Once the baseplate is positioned, the polymerization of the tube/sheath structure is initiated. Further co-localization studies demonstrated that the initial static TssA clusters are then pushed towards the opposite side of the cell by the elongation of the sheath structure. FLIM experiments suggested that the same TssA particle remains associated to the distal extremity of the sheath. This result is also in agreement with the interaction of TssA with both tube (Hcp) and sheath (TssC) components, and with the higher affinity of TssA for these two partners compared to the two baseplate components, suggesting that the recruitment of Hcp and TssBC displaces TssA from the baseplate.

TssA exhibits interesting structural characteristics. Twelve TssA proteins assemble a six-fold symmetry starfish-like particle composed of a central core bearing six elongated 170 Å long arms. The TssA central core, corresponding to the C-terminal domain, has a size and a shape comparable to that of an Hcp hexamer (Extended Data Fig. 7o) and interacts with the N-terminal gp27-like VgrG module and Hcp, two proteins whose structures have been shown to be superimposable<sup>22</sup>, suggesting that TssA recognizes the same fold. The TssA arms interact with TssE and TssC. TssE is the homologue of gp25, a component of the bacteriophage baseplate wedges that assemble around the gp27 hub<sup>1,8</sup>. Therefore the positioning of the TssA central core on VgrG or Hcp allows the arms to contact the outer wedges or sheath rings. Indeed, molecular docking of the TssA electron microscopy volume to the sheath model in the extended conformation shows that the TssA arms are interdigitated with sheath subunits resulting in very complementary shapes and tight contacts (Extended Data Fig. 8a, b). Such an efficient complementarity cannot be modelled with the contracted tail sheath, suggesting that TssA does not bind, or does so much more weakly, to this conformation (Extended Data Fig. 8c).

The TssA C-terminal domain is a hexaflexagon in which the six wedges contact each other by hinge-like helices at the outside of the structure. We hypothesize that large conformational modifications, such as the displacement of these wedges to the exterior, open a large

central lumen. This lumen will have a diameter ( $\sim 90$  Å) sufficient to accommodate a hexameric Hcp ring. We therefore propose a functional model in which TssA controls the coordinated polymerization of the tube and sheath structure (Fig. 3 and Supplementary Video 1): the recruitment of an Hcp ring by TssA will open the lumen allowing the incorporation of this Hcp ring to the growing structure. Then the TssA arms might be involved in the recruitment and proper positioning of the TssBC strands around this Hcp ring, before the insertion of a new Hcp ring, and so on (Fig. 3a). This model implies that the Hcp hexameric unit is incorporated immediately before the TssBC sheath building block, a hypothesis consistent with data showing that in both bacteriophage T4 and T6SS, tube and sheath grow up from the baseplate together, with the tube protein leading and directing sheath assembly<sup>15,26</sup>. However, the current data cannot rule out that a pre-formed Hcp–TssBC complex is recruited to the growing structure (Fig. 3b). In these models, the TssA arms secure the sheath under an extended state by connecting it to the rigid Hcp tube, explaining how the sheath is maintained in a metastable conformation during elongation. In bacteriophages, sheath contraction is initiated at the baseplate and progresses to the head. We therefore propose that TssA remains attached to the distal end during T6SS sheath contraction until the last TssBC row contracts to prevent energy dissipation and permit proper propulsion of the Hcp tube, before being released. This model is in agreement with the docking simulations showing tight contacts of TssA with the extended sheath only (Extended Data Fig. 8a, b). The function of the TssA protein is therefore different from that of the distal tail proteins (Dit) of *Siphoviridae* that prime tube/sheath polymerization but remain attached to the baseplate<sup>27</sup>. It is however closely related to the function of the gp3/gp15 proteins of *Myoviridae*. Although the TssA and gp15 folds are highly divergent, the overall architecture of the central core of TssA is similar in terms of size and diameter to gp15 or gp3 (refs 28, 29) (Extended Data Fig. 8d–f). Notably, the gp3/gp15 proteins, also known as tail terminators, do not only complete tail assembly; pulse-chase studies of bacteriophage T4 biogenesis demonstrated that gp15 is not recruited once the tail tube/sheath has been polymerized but, conversely, that it is assembled on the baseplate before the gp19 tube and gp18 sheath proteins<sup>9</sup>. However, it is found at the opposite extremity of the baseplate once the tail is completed and it stabilizes the sheath<sup>26</sup>. This dynamic is very similar to that observed for TssA, that is, first bound to the baseplate but found at the distal end once the tail is completed. Taken together these data suggest that TssA and gp15 may both prime, control the polymerization, complete and stabilize the T6SS and bacteriophage T4 tails, respectively. Finally, according to this

model, the T6SS tail tube/sheath grows by the incorporation of building blocks at the distal end of the structure, a point that remains to be experimentally addressed. Although the newly incorporated subunits do not transit through the Hcp lumen, this mechanism is similar to that of the flagellar cap complex, which binds to the hook and incorporates new flagellin subunits at the distal end by a rotary mechanism<sup>30</sup>. Defining the TssA conformational modifications that occur during tail elongation will shed light on the molecular mechanism of T6SS tail assembly and how the sheath is maintained in the extended state.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 July 2015; accepted 1 February 2016.**

**Published online 24 February 2016.**

- Leiman, P. G. & Shneider, M. M. Contractile tail machines of bacteriophages. *Adv. Exp. Med. Biol.* **726**, 93–114 (2012).
- Ge, P. *et al.* Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nature Struct. Mol. Biol.* **22**, 377–382 (2015).
- Bönemann, G., Pietrosiuk, A. & Mogk, A. Tubules and donuts: a type VI secretion story. *Mol. Microbiol.* **76**, 815–821 (2010).
- Yang, G., Dowling, A. J., Gerike, U., French-Constant, R. H. & Waterfield, N. R. *Photorhabdus* virulence cassettes confer injectable insecticidal activity against the wax moth. *J. Bacteriol.* **188**, 2254–2261 (2006).
- Shikuma, N. J. *et al.* Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures. *Science* **343**, 529–533 (2014).
- Heymann, J. B. *et al.* Three-dimensional structure of the toxin-delivery particle antifeeding prophage of *Serratia entomophila*. *J. Biol. Chem.* **288**, 25276–25284 (2013).
- Kube, S. & Wendler, P. Structural comparison of contractile nanomachines. *AIMS Biophysics* **2**, 88–115 (2015).
- Leiman, P. G. *et al.* Morphogenesis of the T4 tail and tail fibers. *Virology* **7**, 355 (2010).
- Ferguson, P. L. & Coombs, D. H. Pulse-chase analysis of the *in vivo* assembly of the bacteriophage T4 tail. *J. Mol. Biol.* **297**, 99–117 (2000).
- Rybakova, D. *et al.* Role of antifeeding prophage (Afp) protein Afp16 in terminating the length of the Afp tailocin and stabilizing its sheath. *Mol. Microbiol.* **89**, 702–714 (2013).
- Aschtgen, M. S., Gavioli, M., Dessen, A., Llobès, R. & Cascales, E. The SciZ protein anchors the enteroregulative *Escherichia coli* Type VI secretion system to the cell wall. *Mol. Microbiol.* **75**, 886–899 (2010).
- Felisberto-Rodrigues, C. *et al.* Towards a structural comprehension of bacterial type VI secretion systems: characterization of the TssJ-TssM complex of an *Escherichia coli* pathovar. *PLoS Pathog.* **7**, e1002386 (2011).
- Durand, E. *et al.* Biogenesis and structure of the Type VI secretion membrane core complex. *Nature* **523**, 555–560 (2015).
- Zoued, A. *et al.* Architecture and assembly of the Type VI secretion system. *Biochim. Biophys. Acta* **1843**, 1664–1673 (2014).
- Brunet, Y. R., Hénin, J., Celia, H. & Cascales, E. Type VI secretion and bacteriophage tail tubes share a common assembly pathway. *EMBO Rep.* **15**, 315–321 (2014).
- Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl Acad. Sci. USA* **106**, 4154–4159 (2009).
- Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**, 182–186 (2012).
- English, G., Byron, O., Cianfanelli, F. R., Prescott, A. R. & Coulthurst, S. J. Biochemical analysis of TssK, a core component of the bacterial type VI secretion system, reveals distinct oligomeric states of TssK and identifies a TssK-TssFG subcomplex. *Biochem. J.* **461**, 291–304 (2014).
- Brunet, Y. R., Zoued, A., Boyer, F., Douzi, B. & Cascales, E. The Type VI secretion TssEFGK-VgrG phage-like baseplate is recruited to the TssJLM membrane complex via multiple contacts and serves as assembly platform for tail tube/sheath polymerization. *PLoS Genet.* **11**, e1005545 (2015).
- Basler, M., Ho, B. T. & Mekalanos, J. J. Tit-for-tat: type VI secretion system counterattack during bacterial cell-cell interactions. *Cell* **152**, 884–894 (2013).
- Brunet, Y. R., Espinosa, L., Harchouni, S., Mignot, T. & Cascales, E. Imaging type VI secretion-mediated bacterial killing. *Cell Rep.* **3**, 36–41 (2013).
- Kube, S. *et al.* Structure of the VipA/B type VI secretion complex suggests a contraction-state-specific recycling mechanism. *Cell Rep.* **8**, 20–30 (2014).
- Kudryashev, M. *et al.* Structure of the type VI secretion system contractile sheath. *Cell* **160**, 952–962 (2015).
- Kaptein, N. *et al.* ClpV recycles VipA/VipB tubules and prevents non-productive tubule formation to ensure efficient type VI protein secretion. *Mol. Microbiol.* **87**, 1013–1028 (2013).
- Zoued, A. *et al.* TssK is a trimeric cytoplasmic protein interacting with components of both phage-like and membrane anchoring complexes of the type VI secretion system. *J. Biol. Chem.* **288**, 27031–27041 (2013).
- King, J. Assembly of the tail of bacteriophage T4. *J. Mol. Biol.* **32**, 231–262 (1968).
- Vegge, C. S. *et al.* Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901–1. *J. Bacteriol.* **187**, 4187–4197 (2005).
- Pell, L. G. *et al.* The X-ray crystal structure of the phage  $\lambda$  tail terminator protein reveals the biologically relevant hexameric ring structure and demonstrates a conserved mechanism of tail termination among diverse long-tailed phages. *J. Mol. Biol.* **389**, 938–951 (2009).
- Fokine, A. *et al.* The molecular architecture of the bacteriophage T4 neck. *J. Mol. Biol.* **425**, 1731–1744 (2013).
- Yonekura, K. *et al.* The bacterial flagellar cap as the rotary promoter of flagellin self-assembly. *Science* **290**, 2148–2152 (2000).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was funded by the Centre National de la Recherche Scientifique, the Aix-Marseille Université, and grants from the Agence Nationale de la Recherche to E.C. (ANR-10-JCJC-1303-03), to E.C. and C.C. (ANR-14-CE14-0006-02) and from the Fondation pour la Recherche Médicale to C.C. (FRM DEQ2011-0421282) and supported by the French Infrastructure for Integrated Structural Biology (FRISBI, ANR-10-INSB-05-01). A.Z. and Y.R.B. were supported by doctoral fellowships from the French Ministry of Research. A.Z. and E.D. were supported by end-of-thesis (FDT20140931060) and post-doctoral (SPF20101221116) fellowships from the Fondation pour la Recherche Médicale, respectively. We gratefully acknowledge the Soleil synchrotron radiation facility for beamtime allocation. We thank R. Llobès, J. Sturgis and A. Galinier for constant support, the members of the Cascales, Cambillau, Llobès, Sturgis and Bouveret research groups for helpful discussions, E. Bouveret for providing vectors, protocols and advice for the bacterial two-hybrid assay, C. Bebeacua for preliminary electron microscopy analyses, L. Espinosa for help regarding statistical analyses, R. Lebrun and S. Lignon (proteomic platform, IMM) for mass spectrometry analyses, Y. Cully for the Supplementary Video, O. Uderso, I. Bringer and A. Brun for technical assistance, and J. D. Barras-Elatable for encouragement.

**Author Contributions** A.Z., E.D., C.C. and E.C. designed and conceived the experiments. C.C. and E.C. supervised the execution of the experiments. A.Z., E.D., Y.R.B., S.S., B.D. and M.G. performed the experiments. A.Z. performed the *in vivo* experiments (BACTH, fluorescence microscopy) with the help of Y.R.B., M.G., N.F., L.J. and T.M. E.D. performed the *in vitro* experiments (protein purification and characterization, SAXS, electron microscopy and X-ray analyses) with the help of S.S., P.L. and R.F. B.D. performed the SPR experiments. P.L., R.F., T.M., C.C. and E.C. provided tools. E.C. wrote the paper with contributions of A.Z., E.D. and C.C.

**Author Information** Coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 4Y03 and 4Y05 for TssA<sub>NH2</sub> and TssA<sub>CT</sub> respectively. Electron microscopy map for full-length TssA has been deposited in the Electron Microscopy Data Bank under accession code EMD-3282. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.D. ([edurand@imm.cnrs.fr](mailto:edurand@imm.cnrs.fr)), C.C. ([cambillau@afmb.univ-mrs.fr](mailto:cambillau@afmb.univ-mrs.fr)) or E.C. ([cascales@imm.cnrs.fr](mailto:cascales@imm.cnrs.fr)).

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Bacterial strains, growth conditions and chemicals.** The strains, plasmids and oligonucleotides used in this study are listed in the Supplementary Table. The enteroaggregative *E. coli* EAEC strain 17-2 and its  $\Delta tssA$ ,  $\Delta tssBC$ ,  $\Delta tssE$ ,  $\Delta tssF$ ,  $\Delta tssG$ ,  $\Delta tssK$ ,  $\Delta tssL$ ,  $\Delta tssM$ ,  $\Delta hcp$ ,  $\Delta vgrG$  and *tssB-mCherry* isogenic derivatives were used for this study<sup>11,15,25</sup>. The *E. coli* K-12 DH5 $\alpha$ , W3110, BTH101 and T7 Iq pLys strains were used for cloning steps, co-immunoprecipitation, bacterial two-hybrid and protein purification, respectively. Strains were routinely grown in LB rich medium (or Terrific broth medium for protein purification) or in Sci-1 inducing medium (SIM; M9 minimal medium, glycerol 0.2%, vitamin B1 1  $\mu\text{g ml}^{-1}$ , casaminoacids 100  $\mu\text{g ml}^{-1}$ , LB 10%, supplemented or not with bactoagar 1.5%) with shaking at 37°C<sup>31</sup>. Plasmids were maintained by the addition of ampicillin (100  $\mu\text{g ml}^{-1}$  for *E. coli* K-12, 200  $\mu\text{g ml}^{-1}$  for EAEC), kanamycin (50  $\mu\text{g ml}^{-1}$ ) or chloramphenicol (30  $\mu\text{g ml}^{-1}$ ). Expression of genes from pBAD, pETG20A/pRSF or pASK-IBA vectors was induced at  $A_{600\text{ nm}} \approx 0.6$  with 0.02% of L-arabinose (Sigma-Aldrich) for 45 min, 0.5–1 mM of isopropyl- $\beta$ -D-thio-galactopyranoside (IPTG, Eurobio) for 14 h or 0.02  $\mu\text{g ml}^{-1}$  of anhydrotetracycline (AHT, IBA Technologies) for 45 min, respectively. For BACTH experiments, plates were supplemented with 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-Gal, Eurobio, 40  $\mu\text{g ml}^{-1}$ ). **Strain construction.** The *tssA* gene was deleted into the enteroaggregative *E. coli* 17-2 strain using a modified one-step inactivation procedure<sup>32</sup> as previously described<sup>11</sup> using plasmid pKOBEG<sup>33</sup>. In brief, a kanamycin cassette was amplified from plasmid pKD46 using oligonucleotides carrying 50-nucleotide extensions homologous to regions adjacent to *tssA*. After electroporation of 600 ng of column-purified PCR product, kanamycin-resistant clones were selected and verified by colony-PCR. The kanamycin cassette was then excised using plasmid pCP20 (ref. 32). The deletion of *tssA* was confirmed by colony-PCR. The same procedure was used to introduce the *mCherry*-coding sequence upstream the stop codon of the *tssB* gene (vector *pmCh*-KD4 as template for PCR amplification) or the *sfGFP*-coding sequence downstream the start codon (vector pKD4-*sfGFP* as template) or upstream the stop codon (vector *psfGFP*-KD4 as template) of the *tssA* gene to yield strains producing TssB-mCherry, sfGFP-TssA or TssA-sfGFP from their original chromosomal loci.

**Plasmid construction.** All bacterial two-hybrid plasmids and the plasmid producing the TssJLM membrane core complex (pRSF-TssJ<sup>ST-FL</sup>-TssL<sup>6His</sup>-TssM, pRSF-TssJLM) have been described previously<sup>13,25</sup>. PCR was performed using a Biometra thermocycler using the Q5 (New England Biolabs) or Pfu Turbo (Agilent Technologies) DNA polymerases. Restriction enzymes were purchased from New England Biolabs and used according to the manufacturer's instructions. Custom oligonucleotides were synthesized by Sigma Aldrich and are listed in the Supplementary Table. Enteroaggregative *E. coli* 17-2 chromosomal DNA was used as a template for all PCR. *E. coli* strain DH5 $\alpha$  was used for cloning procedures. All the plasmids (except for pETG20A and pDEST17 derivatives) have been constructed by restriction-free cloning<sup>34</sup> as previously described<sup>25</sup>. In brief, the gene of interest was amplified using oligonucleotides introducing extensions annealing to the target vector. The double-stranded product of the first PCR has then been used as oligonucleotides for a second PCR using the target vector as template. PCR products were then treated with *DpnI* to eliminate template plasmids and transformed into DH5 $\alpha$ -competent cells. For protein purification, the sequences encoding the full-length TssA (residues 1–542), the TssA<sub>Nt</sub> (residues 1–392), the TssA<sub>Nt2</sub> (residues 221–377) and TssA<sub>Ct</sub> (residues 393–542) domains, the N-terminal domain of VgrG (residues 1–490), the full-length TssE or both TssB and TssC were cloned into the pETG-20A (TssA, TssA<sub>Ct</sub>, VgrG<sub>Nt</sub>, TssE) or pDEST17 (TssA<sub>Nt</sub>, TssB, TssC) expression vector (gifts from A. Geerloff, EMBL, Hamburg) according to standard Gateway protocols. Proteins produced from pETG20A derivatives are fused to an N-terminal 6  $\times$  His-tagged thioredoxin (TRX) followed by a cleavage site for the Tobacco etch virus (TEV) protease whereas proteins produced from pDEST17 are fused to an N-terminal 6  $\times$  His tag followed by a TEV protease cleavage site. All constructs have been verified by restriction analyses and DNA sequencing (Eurofins or MWG).

**Bacterial two-hybrid assay (BACTH).** The adenylate cyclase-based bacterial two-hybrid technique<sup>35</sup> was used as previously published<sup>36</sup>. In brief, the proteins to be tested were fused to the isolated T18 and T25 catalytic domains of the *Bordetella* adenylate cyclase. After introduction of the two plasmids producing the fusion proteins into the reporter BTH101 strain, plates were incubated at 30°C for 48 h. Three independent colonies for each transformation were inoculated into 600  $\mu\text{l}$  of LB medium supplemented with ampicillin, kanamycin and IPTG (0.5 mM). After overnight growth at 30°C, 10  $\mu\text{l}$  of each culture were dropped onto LB plates supplemented with ampicillin, kanamycin, IPTG and X-Gal and incubated for 16 h at 30°C. The experiments were done at least in triplicate and a representative result is shown.

**Fluorescence microscopy and image treatment.** Fluorescence microscopy experiments have been performed essentially as described<sup>13,15,21,25</sup>. In brief, cells were grown overnight in LB medium and diluted to an  $A_{600\text{ nm}} \approx 0.04$  into Sci-1 inducing medium (SIM). Exponentially growing cells ( $A_{600\text{ nm}} \approx 0.8$ –1) were harvested, washed in phosphate buffered saline buffer (PBS), resuspended in PBS to an  $A_{600\text{ nm}} \approx 50$ , spotted on a thin pad of 1.5% agarose in PBS, covered with a cover slip and incubated for one hour at 37°C before microscopy acquisition. For each experiment, ten independent fields were manually defined with a motorized stage (Prior Scientific) and stored (x, y, z, PFS-offset) in our custom automation system designed for time-lapse experiments. Fluorescence and phase contrast micrographs were captured every 30 s, using an automated and inverted epifluorescence microscope TE2000-E-PFS (Nikon, France) equipped with a perfect focus system (PFS). PFS automatically maintains focus so that the point of interest within a specimen is always kept in sharp focus at all times despite mechanical or thermal perturbations. Images were recorded with a CoolSNAP HQ 2 (Roper Scientific, Roper Scientific SARL, France) and a 100 $\times$ /1.4 DLL objective. The excitation light was emitted by a 120 W metal halide light. All fluorescence images were acquired with a minimal exposure time to minimize bleaching and phototoxicity effects. The sfGFP images were recorded by using the ET-GFP filter set (Chroma 49002) using an exposure time of 200–400 ms. The mCherry images were recorded by using the ET-mCherry filter set (Chroma 49008) using an exposure time of 100–200 ms. Slight movements of the whole field during the time of the experiment were corrected by registering individual frames using StackReg and Image Stabilizer plugins for ImageJ. sfGFP and mCherry fluorescence channels were adjusted and merged using ImageJ (<http://rsb.info.nih.gov/ij/>). For statistical analyses, fluorescent foci were automatically detected. First, noise and background were reduced using the 'Subtract Background' (20 pixels Rolling Ball) plugin from Fiji<sup>37</sup>. The sfGFP foci were automatically detected by a simple image processing: (1) create a mask of cell surface and dilate; (2) detect the individual cells using the 'Analyse particle' plugin of Fiji; (3) sfGFP foci were identified by the 'Find Maxima' process in Fiji. To avoid false positives, each event was manually controlled in the original data. Box-and-whisker representations of the number of foci per cell were made with R software. *t*-tests were performed in R to statistically compare each population. Kymographs were obtained after background fluorescence subtraction and sectioning using the Kymoreslicewise plug-in under Fiji<sup>37</sup>. Fluorescent foci were detected using a local and sub-pixel resolution maxima detection algorithm and tracked over time with a specifically developed plug-in for ImageJ. The *x* and *y* coordinates were obtained for each fluorescent focus on each frame. The mean square displacement (MSD) was calculated as the distance of the foci from its location at *t* = 0 at each time using R software and plotted over time. For each strain tested, the MSD of at least 10 individual focus trajectories was calculated. For statistical analyses of mobile trajectories, kymograph analyses were performed and the percentage of fixed, mobile with random dynamics and mobile with unidirectional trajectory foci are reported.

**Fluorescence lifetime imaging (FLIM).** FLIM experiments were carried on the same microscope device used for the time-lapse microscopy experiments except with a laser of 488 nm. For each cell a region of interest that corresponds to the size of the laser beam was focused away from TssB-mCherry sheath-labelled sfGFP-TssA for a time of 3 s at a maximum intensity of 100%. The extinction of the complete sfGFP-TssA pool was checked by (i) the absence of recovery of bleached sfGFP-TssA-membrane clusters and (ii) by the overall drop and lack of recovery in intracellular intensity.

**Protein purification.** *E. coli* T7 Iq pLys cells bearing pETG20A or pDEST17 derivatives were grown at 37°C in Terrific Broth to an  $A_{600\text{ nm}} \approx 0.9$  and gene expression was induced with 0.5 mM IPTG for 16 h at 17°C. Cells were harvested, resuspended in Tris-HCl 20 mM pH 8.0, NaCl 150 mM and lysozyme (0.25 mg ml<sup>-1</sup>) and broken by sonication. Soluble proteins were separated from inclusion bodies and cell debris by centrifugation 30 min at 20,000g. The His-tagged fusions were purified using ion metal Ni<sup>2+</sup> affinity chromatography (IMAC) using a 5-ml HisTrap column (GE Healthcare) and eluted with a step gradient of imidazole. The fusion proteins were further digested overnight at 4°C by a hexahistidine-tagged TEV protease using a 1:10 (w/w) protease:protein ratio. The TEV protease and contaminants were retained by a second IMAC and the purified proteins were collected in the flow through. Proteins were further separated on preparative Superdex 200 or Superose 6 gel filtration column (GE Healthcare) equilibrated in Tris-HCl 20 mM pH 8.0, NaCl 150 mM. The fractions containing the purified protein were pooled and concentrated by ultrafiltration using the Amicon technology (Millipore, California, USA). The seleno-methionine (SeMet) derivatives of the N- and C-terminal domains of TssA were produced in minimal medium supplemented with 100 mg l<sup>-1</sup> of lysine, phenylalanine and threonine, 50 mg l<sup>-1</sup> of isoleucine, leucine, valine and seleno-methionine. Gene induction and protein purification were performed as described above.



**Limited proteolysis.** The full-length TssA protein was subjected to Proteinase K limited proteolysis (1:10 protease:protein ratio). The reaction was quenched at different time points by the addition of 1 mM PMSF and further boiling for 10 min at 96 °C. Samples were analysed by SDS–PAGE and Coomassie blue staining. Digested products were identified by Edman N-terminal sequencing and electrospray mass spectrometry (Proteomic platform, Institut de Microbiologie de la Méditerranée, Marseille, France).

**Analytical gel-filtration analysis and MALS/QELS/UV/RI-coupled size-exclusion chromatography.** Size-exclusion chromatography (SEC) was performed on an Alliance 2695 HPLC system (Waters) using KW803 and KW804 columns (Shodex) run in Tris–HCl 20 mM pH 8.0, NaCl 150 mM at 0.5 ml per min. MALS, UV spectrophotometry, QELS and RI were monitored with MiniDawn Treos (Wyatt Technology), a Photo Diode Array 2996 (Waters), a DynaPro (Wyatt Technology) and an Optilab rEX (Wyatt Technology), respectively, as described<sup>12</sup>. Mass and hydrodynamic radius calculation were performed with the ASTRA software (Wyatt Technology) using a  $dn/dc$  value of 0.185 ml g<sup>−1</sup>.

**Surface plasmon resonance analysis.** Steady-state interactions were monitored using a BLAcore T200 at 25 °C<sup>12</sup>. All the buffers were filtered on 0.2 µm membranes before use. The HC200m sensor chip (Xantech) was coated with purified Hcp, VgrG, TssE or TssBC complex, immobilized by amine coupling ( $\Delta RU = 4,000–4,300$ ). A control flow-cell was coated with thioredoxin immobilized by amine coupling at the same concentration ( $\Delta RU = 4,100$ ). Purified TssA, TssA N-terminal and TssA C-terminal domains (six concentrations ranging from 3.125 to 100 µM) were injected and binding traces were recorded in duplicate. The signal from the control flow cell and the buffer response were subtracted from all measurements. The dissociation constants ( $K_d$ ) were estimated using the GraphPad Prism 5.0 software on the basis of the steady state levels of  $\Delta RU$ , directly related to the concentration of the analytes. The  $K_d$  were estimated by plotting on  $x$  axis the different concentration of analytes and the different  $\Delta RU$  at a fixed time (5 s before the end of the injection step) on the  $y$  axis. For  $K_d$  calculation, nonlinear regression fit for xy analysis was used and one site (specific binding) as a model which corresponds to the equation  $y = B_{max} \times x / (K_d + x)$ .

**Co-purification experiments.** Different combinations of plasmids were transformed in BL21(DE3): (i) pRSF–TssJLM + pIBA37(+); (ii) pRSF + pIBA37–<sup>FL</sup>TssA; (iii) pRSF–TssJLM + pIBA37–<sup>FL</sup>TssA; and (iv) pRSF–TssJM + pIBA37–<sup>FL</sup>TssA. Transformed BL21(DE3) cells were grown at 37 °C in 200 ml LB medium supplemented with kanamycin and ampicillin until  $A_{600\text{ nm}} \approx 0.6$  and gene induction was achieved by the addition of IPTG (1 mM) and anhydrotetracycline (0.02 µg ml<sup>−1</sup>) during 15 h at 16 °C. After cell lysis through three passages at the French press, total membranes were isolated as described previously<sup>13</sup>. Membranes were solubilized by the addition of 1% Triton X-100 (Affimatrix). Solubilized membrane fractions were purified on a 1 ml Streptactin column (GE Healthcare). The column was washed with buffer S (HEPES 50 mM pH 7.5, NaCl 50 mM, Triton X-100 0.075%) and bound proteins were eluted with buffer S supplemented with desthiobiotin (2.5 mM) and visualized by Coomassie blue staining and immunoblotting. For electron microscopy (EM) analyses, BL21(DE3) cells producing TssJLM and Flag-tagged TssA were grown and the TssJLM–A complex was purified as described for the TssJLM membrane core complex<sup>13</sup>. After the two consecutive affinity columns (His- and Strep-Trap-HP), the pooled fractions were injected onto a Superose 6 10:300 column equilibrated in HEPES 50 mM pH 7.5, NaCl 50 mM supplemented with 0.025% DM–NPG.

**Electron microscopy observation of the TssJLMA complex.** Nine microlitres of the purified TssJLMA complex ( $\sim 0.01$  mg ml<sup>−1</sup>) were incubated to glow-discharged carbon-coated copper grids (Agar Scientific) for 30 s. After absorption, the sample was blotted, washed with three drops of water and then stained with 2% uranyl acetate. Images were collected on an FEI Tecnai F20 FEG microscope operating at a voltage of 200 kV, equipped with a direct electron detector (Falcon II) at 50,000 magnification.

**Transmission electron microscopy, single particle analysis and image processing.** Nine microlitres of the purified full-length TssA protein ( $\sim 0.01$  mg ml<sup>−1</sup>) were incubated on a glow-discharged carbon-coated copper grid (Agar Scientific) for 30 s. After absorption, the sample was blotted, washed with three drops of water and then stained with 2% uranyl acetate. Images were recorded automatically using the EPU software on a FEG microscope operating at a voltage of 200 kV and a defocus range of 0.6–25 nm, using a FEI Falcon-II detector (Gatan) at a nominal magnification of 50,000, yielding a pixel size of 1.9 Å. A dose rate of 25 electrons per Å<sup>2</sup> per second, and an exposure time of 1 s were used. A total of 100,000 particles were automatically selected from 500 independent images and extracted within boxes of 180 pixels × 180 pixels using EMAN2/BOXER<sup>38</sup>. The CTF was estimated and corrected by phase flipping using EMAN2 (e2ctf). All two- and three-dimensional (2D and 3D) classifications and refinements were performed using RELION 1.3 (refs 39, 40). The automatically selected data set was cleaned up by three rounds of reference-free

2D class averaging, and highly populated classes displaying high-resolution features were conserved and a final data set of 20,000 particles was assembled. An initial 3D-model was generated in EMAN2 using using 30 classes. 3D classification was then performed in Relion with five classes. The particles corresponding to most populated class ( $\sim 18,000$ ) were used for refinement. The Relion auto-refine procedure was used to obtain a final reconstruction at  $\sim 19$  Å resolution after masking and with D6 symmetry imposed. Reported resolutions are based on the gold-standard Fourier shell correlation (FSC) 0.143 criterion; the FSC curve was corrected for the effects of a soft mask on the FSC curve using high-resolution noise substitution (Extended Data Fig. 5a)<sup>41</sup>. All density maps were corrected for the modulation transfer function of the detector and then sharpened by applying a negative B-factor ( $-1000$ ) that was estimated using automated procedures. The electron microscopy map of the EAE TssA full-length protein has been deposited in the Electron Microscopy Data Bank under accession number EMD-3282.

**Small-angle X-ray scattering analysis and *ab initio* 3D shape reconstruction.** Small-angle X-ray scattering (SAXS) analyses were performed at the ID29 beamline (European Synchrotron Radiation Facility, Grenoble, France) at a working energy of 12.5 keV ( $\lambda = 0.931$  Å). Thirty microlitres of protein solution at 1.6, 3.7, 7.1, 9.8 and 14.9 mg ml<sup>−1</sup> in Tris–HCl 20 mM pH 8.0, NaCl 150 mM were loaded by a robotic system into a 2-mm quartz capillary mounted in a vacuum and ten independent 10-s exposures were collected on a Pilatus 6M-F detector placed at a distance of 2.85 m for each protein concentration. Individual frames were processed automatically and independently at the beamline by the data collection software (BsxCUBE), yielding radially averaged normalized intensities as a function of the momentum transfer  $q$ , with  $q = 4\pi \sin(\theta)/\lambda$ , where  $2\theta$  is the total scattering angle and  $\lambda$  is the X-ray wavelength. Data were collected in the range  $q = 0.04–6$  nm<sup>−1</sup>. The ten frames were combined to give the average scattering curve for each measurement. Data points affected by aggregation, possibly induced by radiation damage, were excluded. Scattering from the buffer alone was also measured before and after each sample analysis and the average of these two buffer measures was used for background subtraction using the program PRIMUS<sup>42</sup> from the ATSAS package<sup>43</sup>. PRIMUS was also used to perform Guinier analysis<sup>44</sup> of the low  $q$  data, which provides an estimate of the radius of gyration ( $R_g$ ). Regularized indirect transforms of the scattering data were carried out with the program GNOM<sup>45</sup> to obtain  $P(r)$  functions of interatomic distances. The  $P(r)$  function has a maximum at the most probable intermolecular distance and goes to zero at  $D_{max}$ , the maximum intramolecular distance. The values of  $D_{max}$  were chosen to fit with the experimental data and to have a positive  $P(r)$  function. Three-dimensional bead models that fit with the scattering data were built with the program DAMMIF<sup>46</sup>. Ten independent DAMMIF runs were performed using the scattering profile of TssA, with data extending up to 0.35 nm<sup>−1</sup>, using slow mode settings, assuming P6 symmetry and allowing for a maximum 500 steps to grant convergence. The models resulting from independent runs were superimposed using the DAMAVER suite<sup>47</sup> yielding an initial alignment of structures based on their axes of inertia followed by minimisation of the normalized spatial discrepancy (NSD)<sup>48</sup>. The NSD was therefore computed between a set of ten independent reconstructions, with a range of NSD from 0.678 to 0.815. The aligned structures were then averaged, giving an effective occupancy to each voxel in the model, and filtered at half-maximal occupancy to produce models of the appropriate volume that were used for all subsequent analyses. All the models were similar in terms of agreement with the experimental data, as measured by DAMMIF  $\chi$  parameter and the quality of the fit to the experimental curve (calculated  $\sqrt{\chi} = 1.774$ ). The SAXS data parameters are provided in Extended Data Table 1.

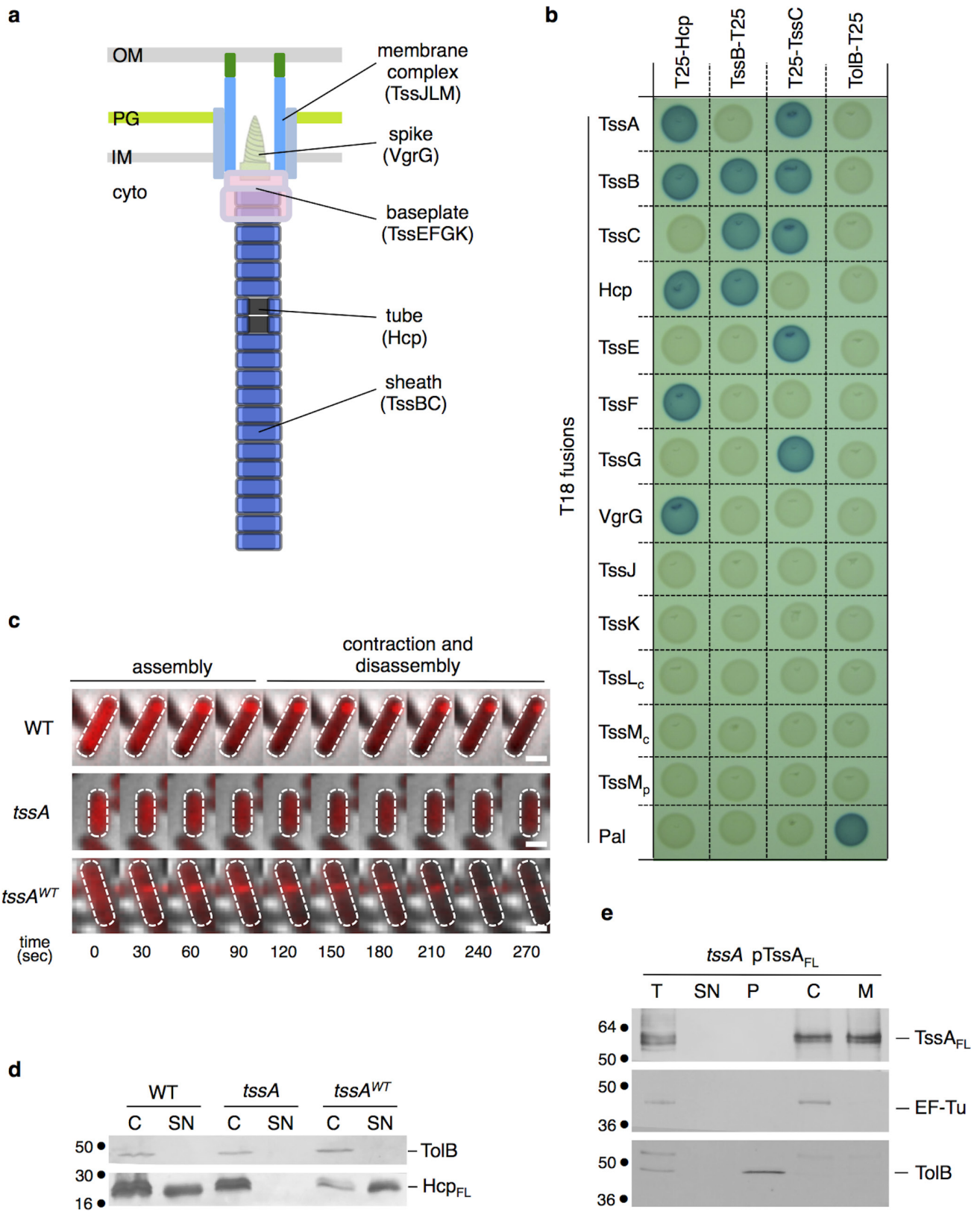
**Crystallization, data collection, processing and refinement.** Seleno-methionine (SeMet)-labelled TssA<sub>Nt</sub> and TssA<sub>Ct</sub> crystallization trials were carried out by the sitting-drop vapour diffusion method in 96-well Greiner crystallization plates at 20 °C, using a nanodrop-dispensing robot (Cartesian Inc.). Crystals of SeMet-labelled TssA<sub>Ct</sub> grew in a few days after mixing 300 nl of protein at 4.7 mg ml<sup>−1</sup> with 100 nl of 20% PEG 8000, 0.2 M calcium acetate, 0.1 M MES pH 6.8. Crystals of SeMet-labelled TssA<sub>Nt</sub> grew in a few days after mixing 300 nl of protein at 4.7 mg ml<sup>−1</sup> with 100 nl of 29% PEG 3350, 0.1 M HEPES pH 7.5. Crystals were cryoprotected with mother liquor supplemented with 20% polyethylene glycol and flash frozen in liquid nitrogen. Data sets were collected at the SOLEIL Proxima 1 beamline (Saint-Aubin, France). After processing the data with XDS<sup>49</sup>, the scaling was performed with SCALA and the structures were solved using the SHELXD program<sup>50</sup>. The structure was refined with AutoBUSTER<sup>51</sup> alternated with model rebuilding using COOT<sup>52</sup>. The final data collection and refinement statistics are provided in Extended Data Table 2. The Ramachandran plots of the TssA<sub>Nt</sub> and TssA<sub>Ct</sub> structures exhibit 90.7/3.3 and 91.8/2.9 residues in the favoured and outlier areas, respectively. Figures were made with PyMOL<sup>53</sup>.

**Tail sheath modelling and TssA docking to contracted and extended sheaths.** The tail sheath modelling was performed using the *Vibrio cholerae* VipAB (TssBC) complex as starting structure<sup>23</sup> (PDB: 3J9G) and the contracted tail sheath structures of *Vibrio cholerae*<sup>23</sup>. To date, however, the molecular structure of the extended

(non-contracted) sheath is not available. In a recent paper, a low-resolution model of the extended VipAB sheath was modelled using the low-resolution EM map of the extended T4 phage tail sheath<sup>22</sup>. By superimposing the VipAB EM map to the gp18 bacteriophage T4 sheath protein structure, gross features of the sheath structure were obtained<sup>22</sup>. A similar approach was applied with Chimera<sup>54</sup> using the VipAB molecular model in the extended T4 phage tail sheath instead of using the low-resolution VipAB EM map, yielding a model similar to that of Kube *et al.*<sup>22</sup>, but with molecular details. The sheath internal channel diameter shrinks from 110 to ~95 Å, and the external diameter from ~290 Å to ~190 Å. The internal diameter of the tail sheath makes it possible to fit stacked Hcp hexamers that are in contact with the tail sheath internal wall. Both extended and contracted tail sheath conformations were used to explore the feasibility of sheath complexes with TssA using its EM map. TssA being at the distal end of the sheath, the polarity of the sheath was taken into consideration. It was suggested that the polarity of T6SS tail sheath is similar to that of bacteriophage T4 and therefore that the VipA (TssB) N-terminal and VipB (TssC) C-terminal helices point to and contact the baseplate<sup>31</sup>. TssA was therefore docked at the opposite extremity of the tail sheath using Chimera<sup>54</sup>. **Miscellaneous.** Hcp release<sup>11,25</sup> and fractionation assays<sup>11,19,25</sup> have been performed as previously described. SDS-polyacrylamide gel electrophoresis was performed using standard protocols. For immunostaining, proteins were transferred onto 0.2-µm nitrocellulose membranes (Amersham Protran), and immunoblots were probed with primary antibodies, and goat secondary antibodies coupled to alkaline phosphatase, and developed in alkaline buffer with 5-bromo-4-chloro-3-indolylphosphate and nitroblue tetrazolium. The anti-TolB polyclonal antibodies are from our laboratory collection, while the anti-Flag (M2 clone, Sigma Aldrich) and anti-EFTu (Roche) monoclonal antibodies and alkaline phosphatase-conjugated goat anti-rabbit or anti-mouse secondary antibodies (Beckman Coulter) have been purchased as indicated.

**Accession numbers.** Coordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 4YO3 and 4YO5 for TssA<sub>Nt</sub> and TssA<sub>Ct</sub>, respectively. Electron microscopy map for full-length TssA has been deposited in the Electron Microscopy Databank (EMDB) under accession code EMD-3282.

31. Brunet, Y. R., Bernard, C. S., Gavioli, M., Lloubès, R. & Cascales, E. An epigenetic switch involving overlapping fur and DNA methylation optimizes expression of a type VI secretion gene cluster. *PLoS Genet.* **7**, e1002205 (2011).
32. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
33. Chaveroche, M. K., Ghigo, J. M. & d'Enfert, C. A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. *Nucleic Acids Res.* **28**, E97 (2000).
34. van den Ent, F. & Löwe, J. RF cloning: a restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods* **67**, 67–74 (2006).
35. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl Acad. Sci. USA* **95**, 5752–5756 (1998).
36. Battesti, A. & Bouveret, E. The bacterial two-hybrid system based on adenylate cyclase reconstitution in *Escherichia coli*. *Methods* **58**, 325–334 (2012).
37. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).
38. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
39. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
40. Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
41. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
42. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
43. Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **39**, 277–286 (2006).
44. Guinier, A. La diffraction des rayons X aux très petits angles; application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys. (Paris)* **12**, 161–237 (1939).
45. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
46. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
47. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
48. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41 (2001).
49. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
50. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
51. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
52. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
53. DeLano, W. L. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
54. Pettersen, E. F. *et al.* UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
55. Gerc, A. J. *et al.* Visualization of the *Serratia* Type VI secretion system reveals unprovoked attacks and dynamic assembly. *Cell Rep.* **12**, 2131–2142 (2015).
56. Douzi, B. *et al.* Crystal structure and self-interaction of the type VI secretion tail-tube protein from enteroaggregative *Escherichia coli*. *PLoS ONE* **9**, e86918 (2014).



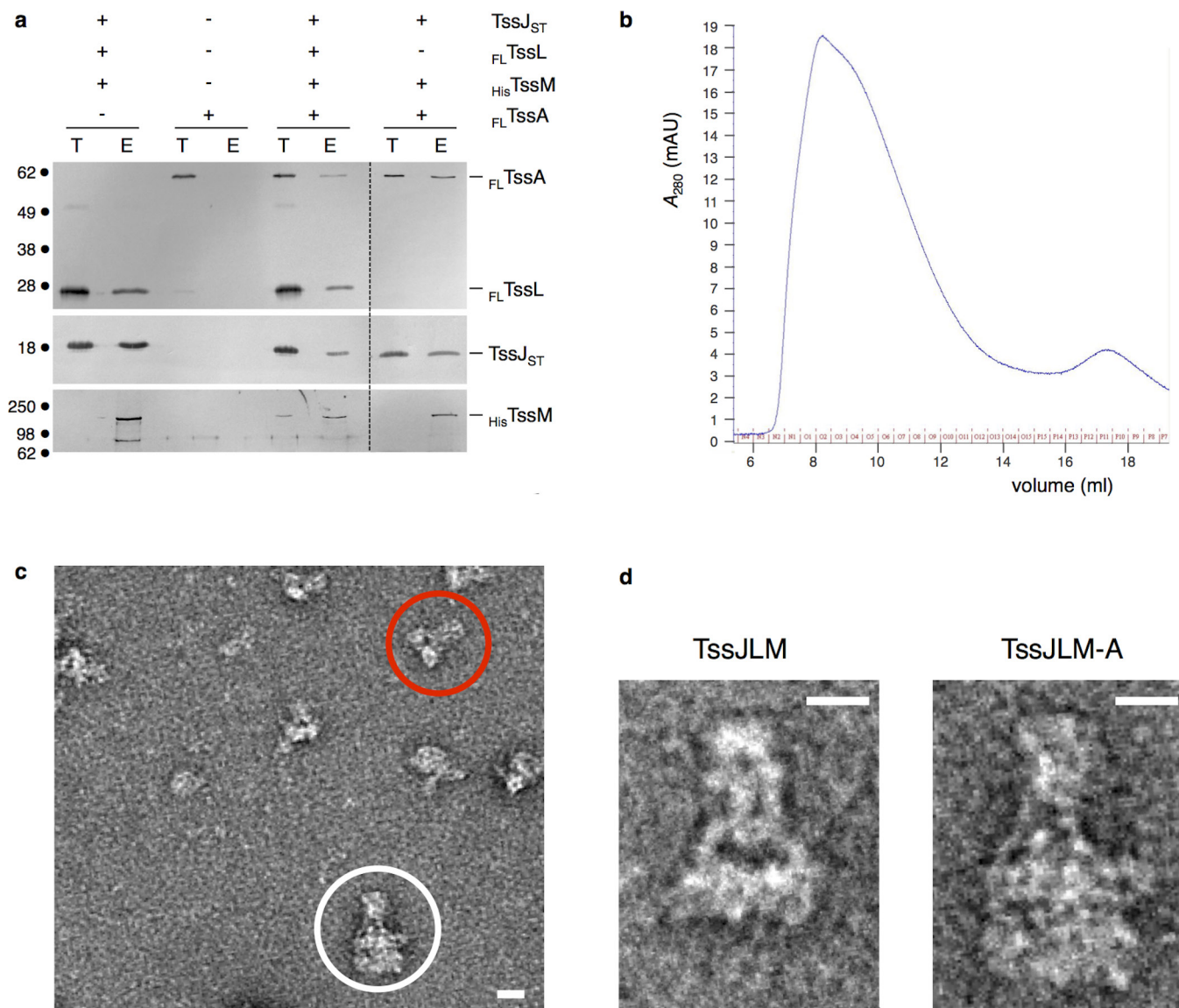
Extended Data Figure 1 | See next page for caption.



# Extended Data Figure 1 | Hcp and TssC interact with TssA, a cytoplasmic protein required for sheath assembly and Hcp release.

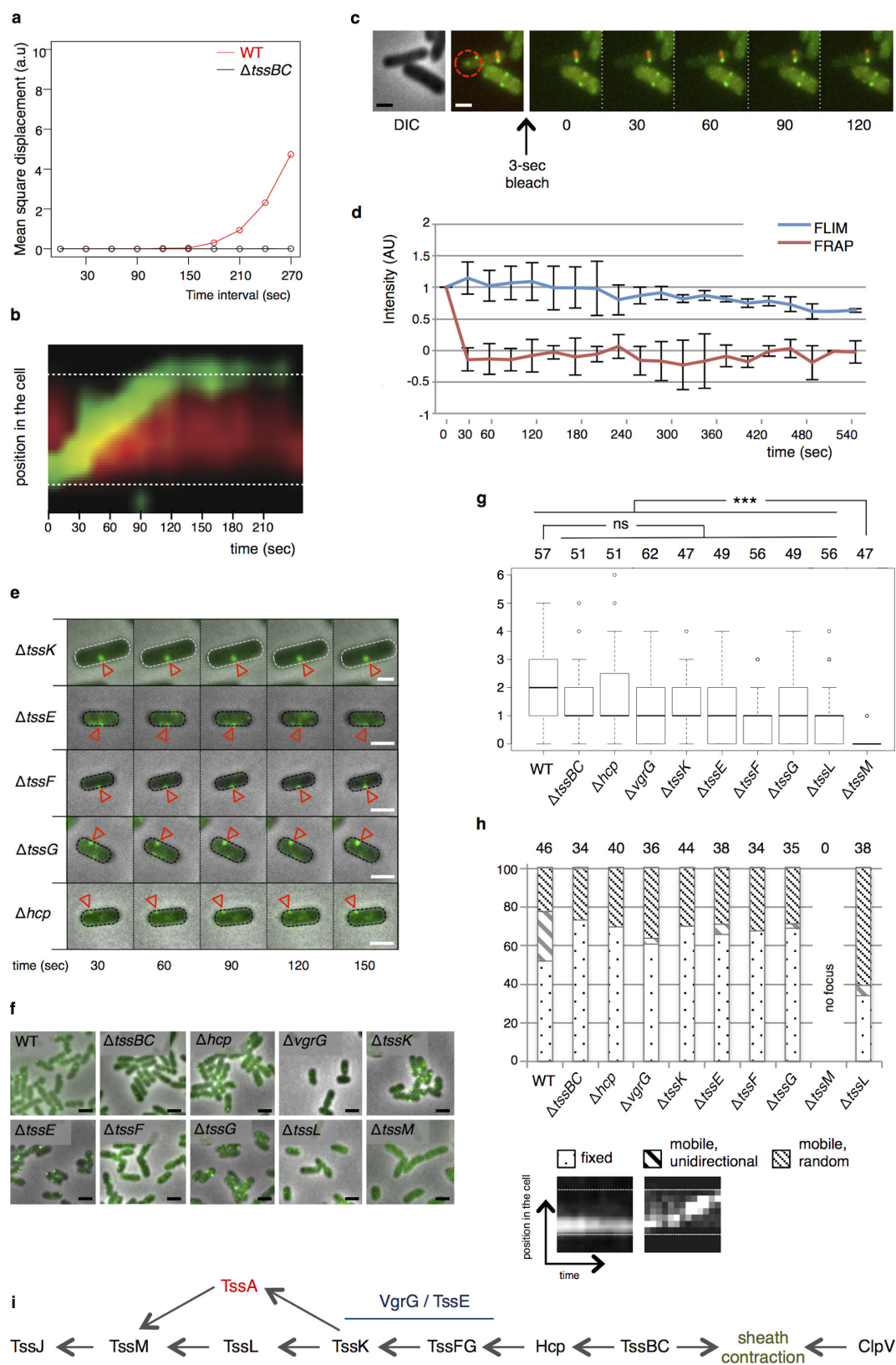
**a.** Schematic representation of the architecture of the bacterial type VI secretion system. The scheme highlights the membrane complex anchoring the tail structure composed of the assembly baseplate, the spike, the tube and the sheath (cyto, cytoplasm; IM, inner membrane; PG, peptidoglycan layer; OM, outer membrane). **b.** Bacterial two-hybrid assay. BTH101 reporter cells producing the indicated proteins or domains (TssLc, cytoplasmic domain of the TssL protein; TssMc and TssMp, cytoplasmic and periplasmic domain of the TssM protein respectively) fused to the T18 or T25 domain of the *Bordetella* adenylate cyclase were spotted on plates supplemented with IPTG and the chromogenic substrate X-Gal. Interaction between the two fusion proteins is attested by the dark blue colour of the colony. The TolB–Pal interaction serves as a positive control. **c.** The absence of TssA prevents T6SS sheath dynamics. Fluorescence microscopy time-lapse recordings showing sheath dynamics using the chromosomally-encoded *tssB-mCherry* fusion in wild-type (WT) (*tssB-mCherry* pBAD33),  $\Delta tssA$  ( $\Delta tssA$  *tssB-mCherry* pBAD33) and complemented  $\Delta tssA$  ( $\Delta tssA$  *tssB-mCherry* pBAD33-*TssA*<sub>VSV-G</sub>) cells. Individual images were taken every 30 s. Assembly and contraction/disassembly events are indicated above the time-lapse images. The scale

bars are 1  $\mu$ m. **d.** The absence of TssA prevents Hcp release. Hcp release was assessed by separating whole cells (C) and supernatant (SN) fractions from 17-2 (WT),  $\Delta tssA$  ( $\Delta tssA$  pBAD33, *tssA*) and complemented  $\Delta tssA$  ( $\Delta tssA$  pBAD33-*TssA*<sub>VSV-G</sub>, *tssA*<sup>WT</sup>) cells producing Flag-epitope-tagged Hcp. A total of  $1 \times 10^8$  cells and the TCA-precipitated material from the supernatant of  $5 \times 10^8$  cells were analysed by western blot using anti-Flag monoclonal antibody (lower panel) and anti-TolB polyclonal antibodies as a lysis control (upper panel). The molecular weight markers (in kDa) are indicated on the left. The uncropped scans of the western blots are provided in the Supplementary Figure. **e.** TssA co-fractionates with cytoplasmic and membrane proteins. A fractionation procedure was applied to EAEC  $\Delta tssA$  cells producing Flag-tagged TssA. Whole cells (T) were fractionated to isolate the supernatant (SN), periplasmic (P), cytoplasmic (C) and total membrane (M) fractions. Extracts from  $10^9$  (T) or  $2 \times 10^9$  (SN, P, C, M) cells were separated by SDS–PAGE and immunodetected with anti-Flag monoclonal (TssA), anti-EF-Tu (cytoplasmic marker) and TolB (periplasmic marker) antibodies. The molecular weight markers (in kDa) are indicated on the left. The uncropped scans of the western blots are provided in Supplementary Figure 1.



**Extended Data Figure 2 | Purification and negative-stain electron microscopy analyses of the TssJLM–TssA complex.** **a**, TssA interacts with the TssJLM complex. The total solubilized membrane extract (T) of  $4 \times 10^9$  cells producing the indicated proteins was subjected to affinity chromatography using streptactin resin. Bound proteins (E) were separated by SDS–PAGE and immunodetected with anti-Flag (TssA and TssL), anti-strep tag (TssJ) and anti-5×His (TssM) monoclonal antibodies. The molecular weight markers are indicated on the left. The uncropped details of the western blots are provided in Supplementary Figure 1.

**b**, Superose 6 10/300 gel-filtration profile of the purified TssJLM–TssA complex. The asymmetry of the peak probably reflects the co-purification of different complexes or the dissociation of TssA from the TssJLM complex. **c**, Examples of representative raw particles observed for the purified TssJLM–TssA complex sample using negative-stain electron microscopy. A typical TssJLM complex is shown in red (number of particles observed  $n = 240$ ) whereas a TssA-bound TssJLM complex is shown in white ( $n = 95$ ). Scale bar is 10 nm. **d**, Magnification of the two complexes shown in **c**. Scale bars, 10 nm.

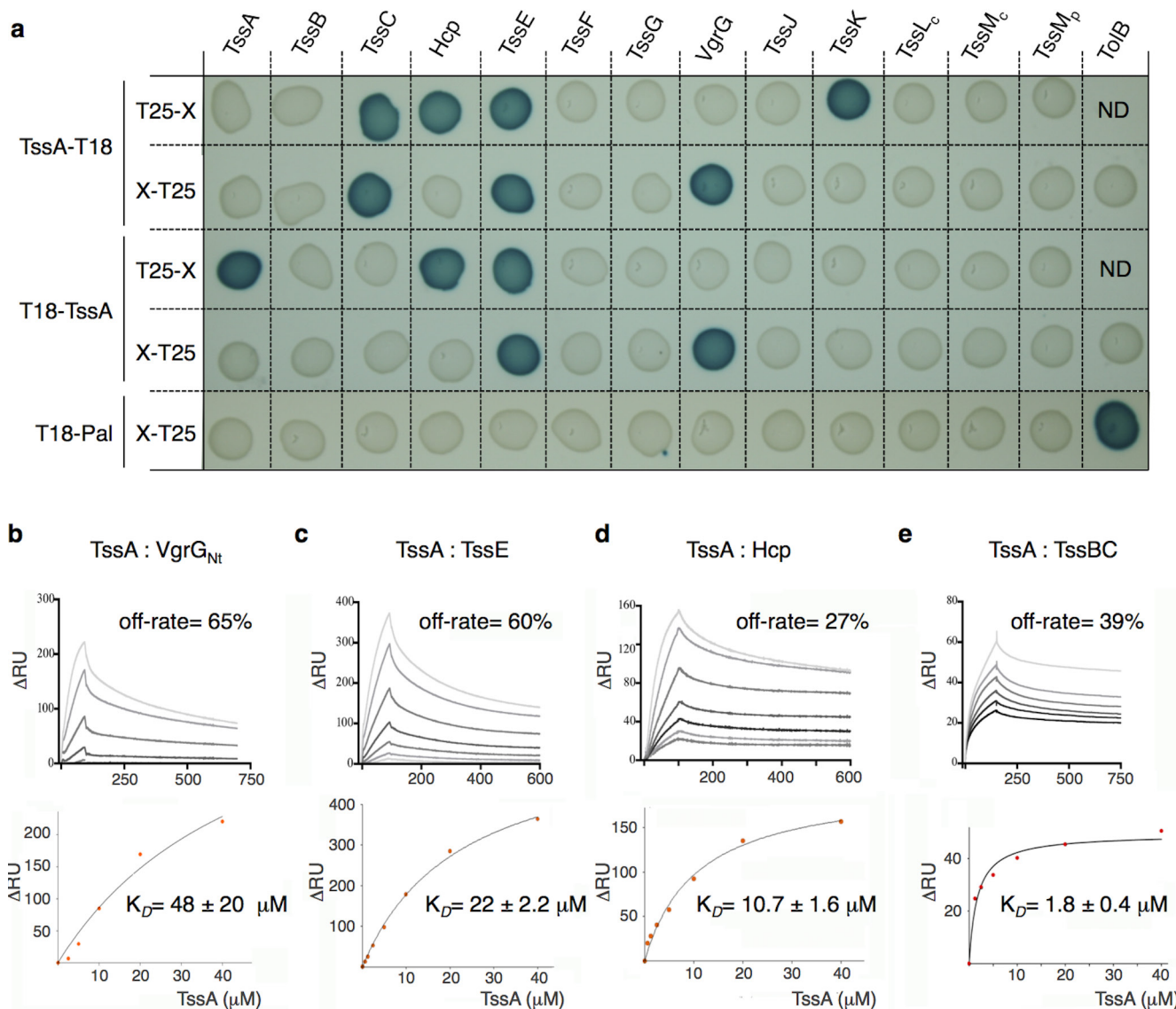


Extended Data Figure 3 | See next page for caption.



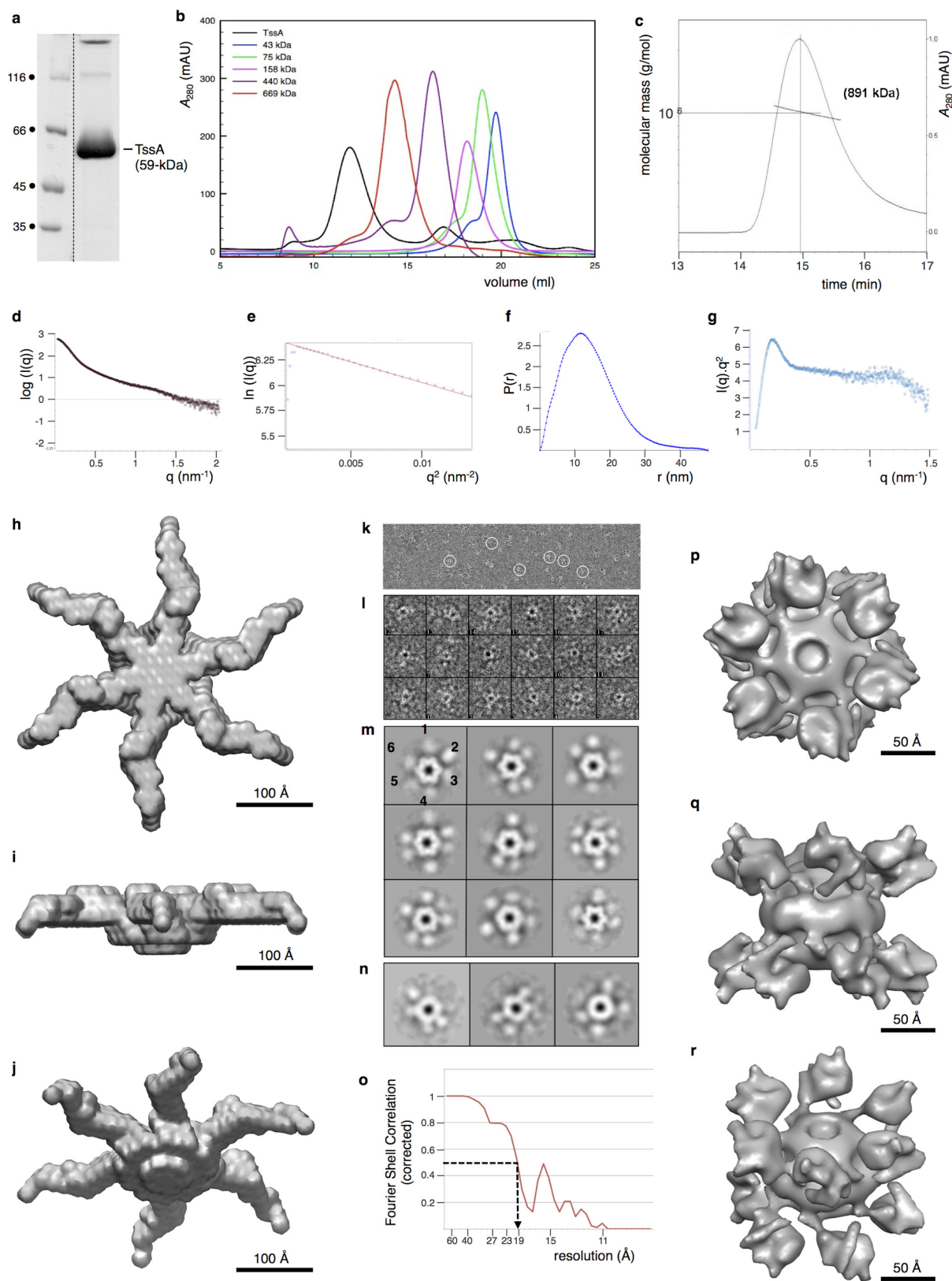
**Extended Data Figure 3 | TssA localization and dynamics.** **a**, Mean square displacement (MSD; in arbitrary units (a.u.)) of a representative sfGFP–TssA clusters in a wild-type strain (red line) or its  $\Delta tssBC$  isogenic derivative (black line) were measured by sub-pixel tracking of fluorescent foci and plotted over time (in sec). **b**, Kymographic analysis reporting representative sfGFP–TssA (green) and TssB–mCherry (red) positions within the cell as a function of time. **c**, Representative fluorescence lifetime imaging microscopy (FLIM) of sfGFP–TssA clusters in the sfGFP–TssA/TssB–mCherry strain. A membrane-associated sfGFP–TssA cluster was chosen to define the bleached area (red circle). The laser (488 nm) was set to maximum power and focused for 3 s to ensure complete bleaching of the GFP diffusible pool. Images were taken every 30 s to follow recovery dynamics. The scale bar is 1  $\mu\text{m}$ . **d**, Quantification of sfGFP–TssA fluorescence dynamics over time after bleaching. The dynamics of fluorescence intensity is shown over time for  $n = 10$  independent sfGFP–TssA foci after FLIM (blue line). The fluorescence intensity of the bleached focus was also followed over time (FRAP, red line). As a control for laser focusing and intensity, membrane-associated clusters were systematically bleached in these experiments and showed no recovery suggesting the total intracellular sfGFP–TssA has been bleached by the laser. **e**, Representative fluorescence microscopy time-lapse recordings of the indicated  $\Delta tssK$ ,  $\Delta tssE$ ,  $\Delta tssF$ ,  $\Delta tssG$  or  $\Delta hcp$  cells producing sfGFP–TssA. Individual images were taken every 30 s. Red arrowheads indicate

the localizations of TssA foci. The scale bar is 1  $\mu\text{m}$ . **f**, Representative large fields of fluorescence microscopy analyses showing localization of sfGFP–TssA in the indicated strains. The scale bars are 1  $\mu\text{m}$ . **g**, Box-and-whisker plots of the measured number of sfGFP–TssA foci per cell for each indicated strain. The lower and upper boundaries of the boxes correspond to the 25% and 75% percentiles respectively. The black bold horizontal bar represents the median values for each strain and the whiskers represent the 10% and 90% percentiles. Outliers are shown as open circles. A Student's  $t$ -test was used to report significant differences (ns, not significant; \*\*\* $P < 0.0001$ ). The number of cells studied per strain ( $n$ ) is indicated on top. **h**, Statistical analyses of sfGFP–TssA dynamics. sfGFP–TssA dynamics were categorized as 'fixed', 'mobile with unidirectional trajectory' and 'mobile with random dynamics' and the number of sfGFP–TssA ( $n$ , on top) foci in each category is represented as a percentage for each indicated strain. Kymographs for the two first categories are shown at the bottom of the panel. **i**, Schematic representation of the assembly pathway of the T6SS based on this study and available data<sup>13,15,19,24,25,55</sup>. The biogenesis starts with the initial positioning of the TssJ outer membrane lipoprotein and the sequential recruitment of the indicated subunits (from left to right). The recruitment of TssA is dependent on TssM, and that of TssK is dependent on both TssL and TssA. The exact positions of VgrG and TssE (blue) in the pathway are not known but these two subunits are not required for TssA recruitment but necessary for Hcp and TssBC polymerization.



**Extended Data Figure 4 | TssA interacts with tail and baseplate components.** **a**, TssA interaction network identified by bacterial two-hybrid analysis (see legend to Extended Data Fig. 1b). **b–e**, Surface plasmon resonance interaction study of TssA with its partners identified by BACTH. Sensorgrams (variation of plasmon resonance in arbitrary units ( $\Delta\text{RU}$ ) as a function of reaction time (in sec)) were recorded upon injection of the purified native TssA protein (concentrations of 3.125

(dark grey), 6.25, 12.5, 25, 50 and 100 (light grey)  $\mu\text{M}$ ) on HC200m chips coated with the purified N-terminal domain of VgrG (**b**), purified TssE (**c**), Hcp (**d**) or TssBC complex (**e**) (upper panels). The graph reporting  $\Delta\text{RU}$  as a function of the TssA concentration (lower panel) was used to estimate the indicated apparent dissociation constants ( $K_D$ ). Off-rates (percentage of dissociation 400 s after ligand injection) are indicated.

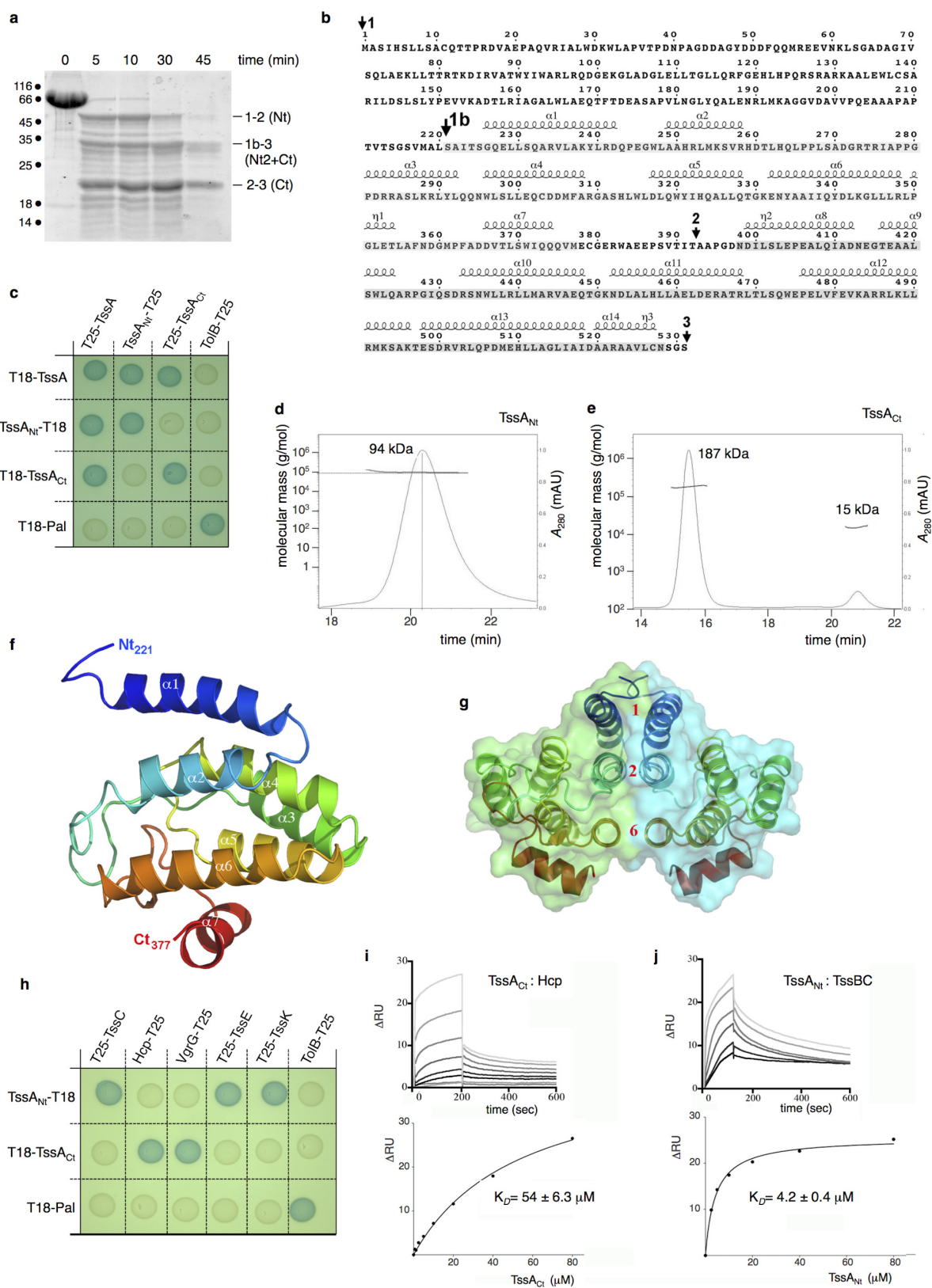


Extended Data Figure 5 | See next page for caption.



**Extended Data Figure 5 | TssA oligomerization and SAXS and EM structural models.** **a**, 10  $\mu$ g of purified TssA were analysed by SDS–PAGE and Coomassie blue staining. The molecular weight markers (in kDa) are indicated on the left, and TssA and its theoretical size are indicated on the right. **b**, Superose 6 10/300 gel filtration profile of purified TssA (black line) and protein markers of known size (coloured lines). **c**, MALS/QELS/UV/RI analysis of purified TssA. The molecular mass of the TssA complex is indicated. **d–j**, Low-resolution SAXS model of TssA. **d**, Experimental scattering data calculated from an *ab initio* model of TssA. The square root  $\chi$  value of the ‘best representative’ model is indicated. **e**, Representation of the Guinier plot calculated from the experimental curve. **f**, Pair distance distribution. **g**, Kratky plot representative of a multi-domain protein with flexible linkers. **h–j**, SAXS envelope of the ‘best representative’ model of TssA, with top (**h**), side (**i**) and tilted (**j**) views. The scale bar is 100 Å.

**k–r**, Low-resolution EM model of TssA. **k**, Representative micrograph of the data set used for image processing. White circles indicate isolated TssA dodecamers. **l**, Representative selected TssA particles. **m, n**, Gallery of representative top (**m**) and side (**n**) class averages generated after reference-free 2D classification using Relion<sup>39</sup>. **o**, Fourier shell correlation (FSC) curve of the TssA reconstruction. The gold standard FSC curve was calculated in Relion<sup>39</sup> using the masked reconstruction of TssA. **p–r**, Top (**p**), side (**q**) and tilted (**r**) views of the three-dimensional reconstruction model of the TssA dodecamer obtained by electron microscopy (accession number: EMD-3282). The scale bar is 50 Å. Whereas the SAXS model allows to better visualize the arm length compared to the EM reconstruction, its low resolution impairs the visual separation of the dimeric arms.



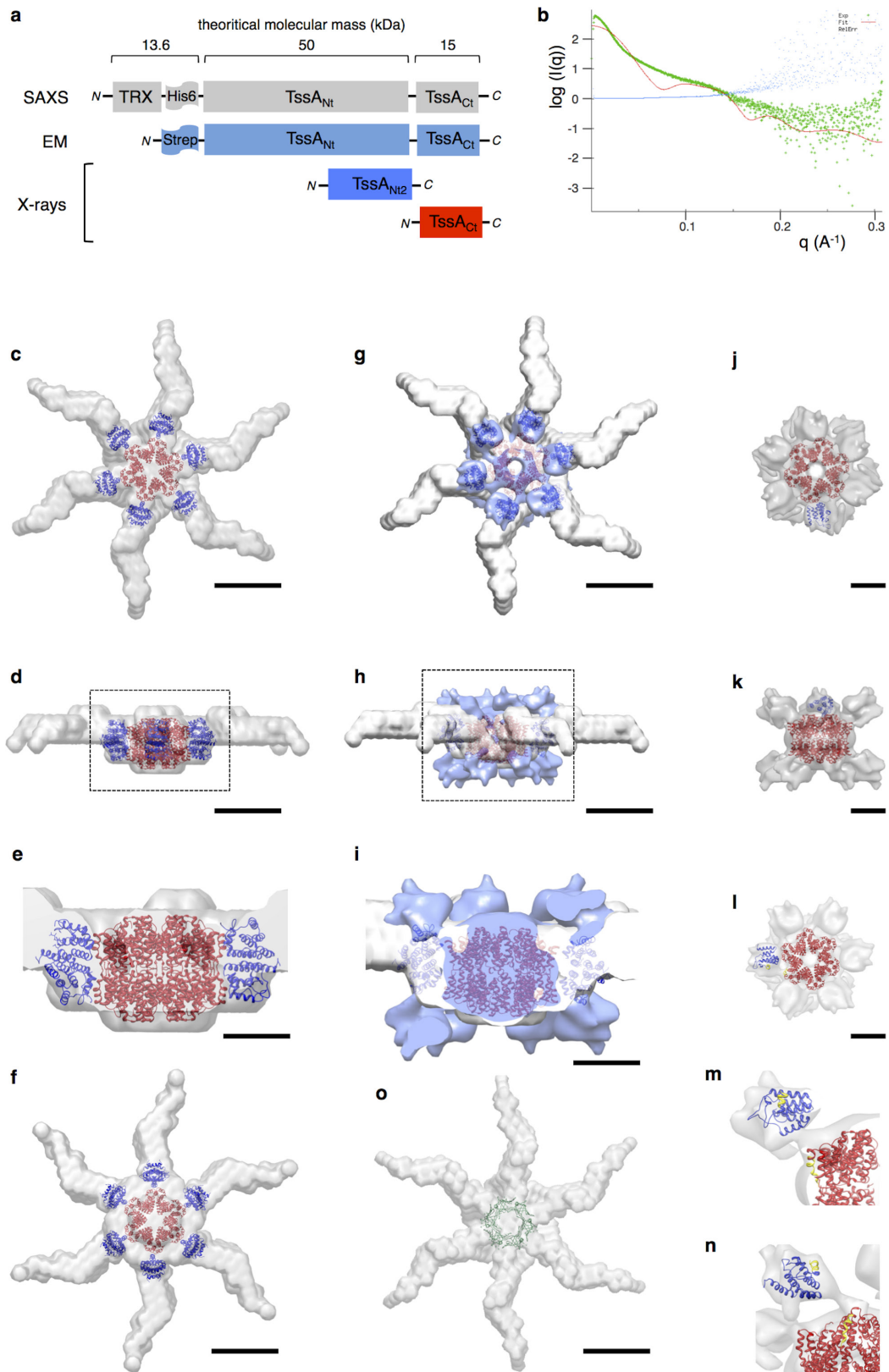
Extended Data Figure 6 | See next page for caption.

**Extended Data Figure 6 | Identification, oligomerization and**

**interaction analysis of TssA domains.** **a**, Limited proteolysis of purified TssA. The purified full-length TssA protein (first lane) was submitted to proteinase K limited proteolysis for the time indicated on top of each lane and analysed by SDS–PAGE and Coomassie blue staining. Stable fragments are indicated on the right with their boundaries (numbers identified in the sequence in **b**) and the corresponding fragment. The uncropped scan of the Coomassie blue stained gel is provided in Supplementary Figure 1. **b**, TssA protein sequence. The localization of the boundaries of the stable fragments obtained after Proteinase K limited proteolysis and electrospray mass spectrometry analyses are arrowed. The secondary structures observed in the crystal structures (Fig. 2a and Extended Data Fig. 6f, g) are indicated on top of the corresponding sequence. **c**, Bacterial two-hybrid analysis of TssA<sub>Nt</sub> and TssA<sub>Ct</sub> interactions (see legend to Extended Data Fig. 1b). **d**, **e**, MALS/QELS/UV/RI analysis of the purified TssA<sub>Nt</sub> (**d**) and TssA<sub>Ct</sub> (**e**) fragments. **f**, **g**, X-ray

structure of the TssA<sub>Nt2</sub> domain (PDB: 4YO3). The rainbow coloured ribbon representation of the TssA<sub>Nt</sub> monomer is shown (**f**, consecutive  $\alpha$ -helices numbered  $\alpha$ 1 to  $\alpha$ 7) whereas the dimeric structure (**g**) highlights the helices at the interface ( $\alpha$ 1,  $\alpha$ 2 and  $\alpha$ 6). **h**, The TssA central core interacts with Hcp and VgrG whereas the TssA arms interact with TssE and TssC. Bacterial two-hybrid analysis of TssA<sub>Nt</sub> and TssA<sub>Ct</sub> interactions (see legend to Extended Data Fig. 1b). **i**, **j**, Surface plasmon resonance interaction study of the purified TssA<sub>Ct</sub> (**i**) or TssA<sub>Nt</sub> (**j**) domains with the Hcp protein (**i**) or the TssBC complex (**j**). Sensorgrams (variation of plasmon resonance in arbitrary unit ( $\Delta$ RU) as a function of reaction time (in seconds)) were recorded upon injection of the purified TssA C-terminal (**i**) or TssA<sub>Nt</sub> (**j**) domains (concentrations of 3.125 (dark grey), 6.25, 12.5, 25, 50 and 100 (light grey)  $\mu$ M) on HC200m chips coated with the purified Hcp protein (**i**) or the TssBC complex (**j**) (upper panels). The graph reporting  $\Delta$ RU as a function of the TssA domain concentration (lower panel) was used to estimate the indicated apparent dissociation constants ( $K_d$ ).





Extended Data Figure 7 | See next page for caption.

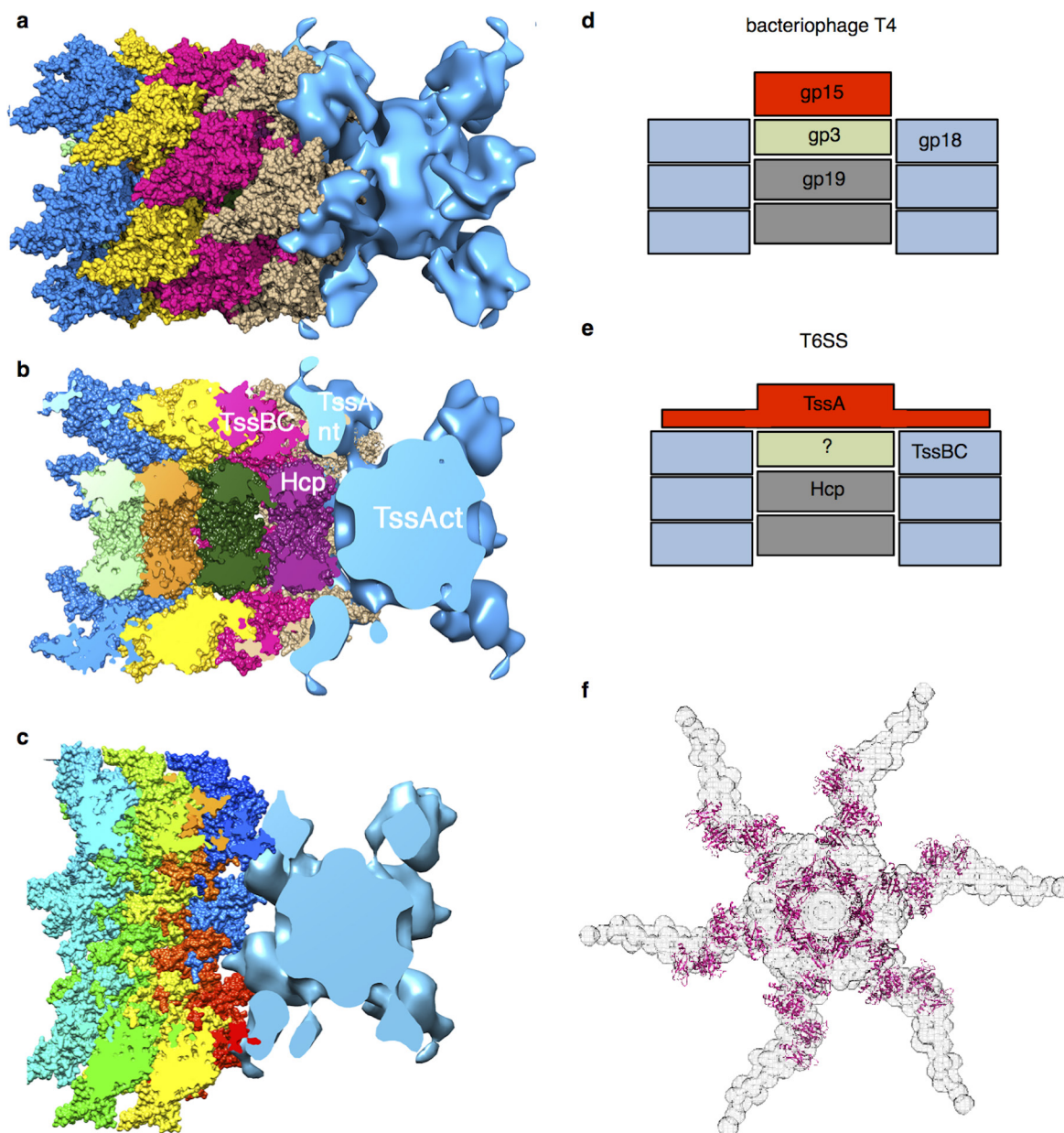
**Extended Data Figure 7 | Comparison of the SAXS, EM and X-ray structures of TssA.** **a**, Schematic representation and colour code of the constructs used for SAXS (grey), electron microscopy (light blue) and X-ray (TssA<sub>Nt</sub>, dark blue; TssA<sub>Ct</sub>, red) analyses. The epitopes and theoretical molecular masses of the domains are indicated.

TRX, thioredoxine; N, N terminus; C, C terminus. **b**, Fit between the experimental data (green dots) and the calculated scattering curves for TssA<sub>Nt2</sub> and TssA<sub>Ct</sub> generated by CRYSOLO (red line). **c–f**, SAXS/X-ray comparison. Top (**c**), side (**d**) and bottom (**f**) views of the fitting of TssA<sub>Nt2</sub> (blue ribbon) and TssA<sub>Ct</sub> (red ribbon) X-ray structures into the TssA SAXS envelope (transparent grey surface). Scale bars are 10 nm. **e**, Magnification of a cut-away section of the fitting shown in **d**. Scale bar is 5 nm.

**g–i**, SAXS/EM/X-ray comparison. Top (**g**) and side (**h**) views of the superimposition of SAXS (grey surface), EM (transparent light-blue

surface) and X-ray structures of TssA. Scale bars are 10 nm. **i**, Magnification of a cut-away section of the superimposition shown in **h**. **j–n**, EM/X-ray comparison. Top (**j**), side (**k**) and bottom (**l**) views of the fitting of TssA<sub>Nt2</sub> (blue ribbon) and TssA<sub>Ct</sub> (red ribbon) X-ray structures into the TssA EM envelope (transparent grey surface). Scale bars are 10 nm.

**m, n**, Magnifications of the top and bottom views of the docking of the TssA domain X-ray structures into the TssA EM map highlighting the interface between the TssA central core (TssA<sub>Ct</sub>, red ribbon) and arms (TssA<sub>Nt2</sub>, blue ribbon). The C-terminal helix of TssA<sub>Nt2</sub> (ends at position 377) and N-terminal helix of TssA<sub>Ct</sub> (starts at position 395) are shown in yellow. **o**, Top view of the fitting of the X-ray structure of EAEC Hcp (green ribbon, PDB 4HKH<sup>56</sup>) into the TssA SAXS envelope (grey surface). The scale bar is 10 nm.



**Extended Data Figure 8 | Models of tail-sheath TssA complexes and comparison between the bacteriophage T4 gp15 and T6SS TssA subunits. a, b**, Surface (a) and cross-section (b) views of the complex of TssA (EM map, blue) with the extended tail sheath model (the last four rows shown in different colours). In the cut-away section, four stacks of Hcp rings are visible. As shown by bacterial two-hybrid and SPR analyses, Hcp contacts the TssA<sub>CT</sub> central core whereas TssBC contacts the TssA<sub>NT</sub> arms. The TssA arms fit between the TssBC monomers of the last row. **c**, Surface view of the complex of TssA (EM map, blue) with the contracted tail sheath model (the last four rows shown in different colours), highlighting the loose packing between TssA and the tail sheath

in this conformation, suggesting that TssA might dissociate after sheath contraction. **d–f**, Comparison between the bacteriophage T4 gp15 and T6SS TssA subunits. **d, e**, Schematic representations of the bacteriophage T4 tail distal end comprising the gp19 tube (grey) and gp18 sheath (blue) proteins and the gp3 (green) and gp15 (red) neck proteins (**d**) and the T6SS tail distal end comprising the Hcp tube (grey) and TssBC sheath (blue) proteins and the TssA dodecamer (red) (**e**). The possibility that a functional homologue of bacteriophage T4 gp3 exists is shown by the question mark. **f**, Fitting of the model of the gp15 structure in complex with the last row of the gp18 sheath (in purple)<sup>29</sup> in the TssA SAXS envelope (grey surface).



Extended Data Table 1 | SAXS parameters of TRX-TssA

Facilities and parameters	Settings and values
<b>Data Collection Parameters</b>	
Beam line	ESRF (Grenoble, France) BM29
Wavelength (Å)	0.992
Detector	Pilatus 1M
$q$ range (nm <sup>-1</sup> )	0.028-4.525
Exposure time (s)	1 (10 × 10 sec)
Concentration range (mg/ml)	1.6-14.9
Temperature (°C)	20
<b>Structural Parameters</b>	
$R_g$ from Guinier fitting (Å)	107.1 ± 0.25
$I(0)$ from Guinier fitting	615.49 ± 1.16
$R_g$ from GNOM (Å)	114.5 ± 0.055
$I(0)$ from GNOM	629.7 ± 1.54
$D_{max}$ (Å)	420 ± 0.093
$V_{porod}$ from PRIMUS (Å <sup>3</sup> )	3081.4 × 10 <sup>3</sup>
$MW^{Pred}$ (kDa) <sup>a</sup>	891
$MW^{SEC-MALS}$ (kDa) <sup>b</sup>	952
$MW^{SAXS}$ (kDa) <sup>c</sup>	629
NSD of DAMMIF models	0.678-0.815
<b>Software Employed</b>	
Primary Data Processing	PRIMUS
$P(r)$	GNOM
<i>Ab initio</i> Shape Analysis	DAMMIF
Validation and averaging	DAMAVR
SAXS Profile computation	CRYSOL
Molecular Visualization	CHIMERA; PyMol

<sup>a</sup> $MW^{pred}$ , theoretical mass of the TRX-TssA protein.<sup>b</sup> $MW^{SEC-MALS}$ , mass measured by SEC-MALS experiments.<sup>c</sup> $MW^{SAXS}$ , mass calculated from the  $I(0)$ .

Extended Data Table 2 | Data collection, phasing and refinement statistics for SAD (SeMet) structures

	TssA <sub>Nt2</sub> (PDB: 4YO3) <sup>a</sup>	TssA <sub>Ct</sub> (PDB: 4YO5) <sup>a</sup>
<b>Data collection</b>		
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub>
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	134.1, 178.0, 106.8	112.3, 90.6, 120.0
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 102, 90
	<i>Peak</i>	<i>Peak</i>
Wavelength	0.987	0.987
Resolution (Å)*	28.4-3.37 (3.47-3.37)	46.2-3.35 (3.45-3.35)
<i>R</i> <sub>merge</sub> *	0.15 (0.72)	0.12 (0.76)
<i>I</i> / $\sigma$ <i>I</i> *	11.4 (2.3)	8.2 (2.0)
Completeness (%)*	99.7 (99.9)	99 (99)
Redundancy*	7.8 (7.9)	4 (4)
<b>Refinement</b>		
Resolution (Å)*	28.4-3.37 (3.47-3.37)	46.2-3.35 (3.45-3.35)
No. reflections*	36772 (2999)	33971 (2906)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> *	25.1/26.1 (27.6/28.5)	22.6/26.0 (25.1/27.1)
No. atoms		
Protein	14496	12298
Ligand/ion	0	0
Water	0	0
B-factors		
Protein	111	103
Ligand/ion		
Water		
R.m.s deviations		
Bond lengths (Å)	0.01	0.01
Bond angles (°)	1.18	1.17

<sup>a</sup>One crystal used for data collection.

\*Highest resolution shell is shown in parentheses.

# Structural basis of outer membrane protein insertion by the BAM complex

Yinghong Gu<sup>1\*</sup>, Huanyu Li<sup>1\*</sup>, Haohao Dong<sup>1\*</sup>, Yi Zeng<sup>1\*</sup>, Zhengyu Zhang<sup>1\*</sup>, Neil G. Paterson<sup>2\*</sup>, Phillip J. Stansfeld<sup>3\*</sup>, Zhongshan Wang<sup>1,4,5</sup>, Yizheng Zhang<sup>5</sup>, Wenjian Wang<sup>6</sup> & Changjiang Dong<sup>1</sup>

All Gram-negative bacteria, mitochondria and chloroplasts have outer membrane proteins (OMPs) that perform many fundamental biological processes. The OMPs in Gram-negative bacteria are inserted and folded into the outer membrane by the  $\beta$ -barrel assembly machinery (BAM). The mechanism involved is poorly understood, owing to the absence of a structure of the entire BAM complex. Here we report two crystal structures of the *Escherichia coli* BAM complex in two distinct states: an inward-open state and a lateral-open state. Our structures reveal that the five polypeptide transport-associated domains of BamA form a ring architecture with four associated lipoproteins, BamB–BamE, in the periplasm. Our structural, functional studies and molecular dynamics simulations indicate that these subunits rotate with respect to the integral membrane  $\beta$ -barrel of BamA to induce movement of the  $\beta$ -strands of the barrel and promote insertion of the nascent OMP.

OMPs have important roles in Gram-negative bacteria, mitochondria and chloroplasts in nutrition transport, protein import, secretion and other fundamental biological processes<sup>1–3</sup>. Dysfunction of mitochondria outer membrane proteins is linked to disorders such as diabetes, Parkinson disease and other neurodegenerative diseases<sup>4,5</sup>. The OMPs are inserted and folded correctly into the outer membrane (OM) by the conserved OMP85 family proteins<sup>6–8</sup>, suggesting that similar insertion mechanisms may be used in Gram-negative bacteria, mitochondria and chloroplasts.

In Gram-negative bacteria, OMPs are synthesized in the cytoplasm, and are transported across the inner membrane by SecYEG into the periplasm<sup>8,9</sup>. The 17-kilodalton (kDa) protein (Sbp) and the survival factor A (SurA) chaperones escort the unfolded OMPs across the periplasm to the BAM, which is responsible for insertion and assembly of OMPs into the OM<sup>10–12</sup>. In *E. coli*, the BAM complex consists of BamA and four lipoprotein subunits, BamB, BamC, BamD and BamE. BamA is comprised of five amino-terminal polypeptide transport-associated (POTRA) domains and a carboxy-terminal OMP transmembrane barrel, while the four lipoproteins are affixed to the membrane by N-terminal lipid-modified cysteines. Of these subunits, BamA and BamD are essential<sup>3,6</sup>. One copy of each of these five proteins is required to form the BAM complex with an approximate molecular mass of 200 kDa (Extended Data Fig. 1). *In vitro* reconstitution of the *E. coli* BAM complex and functional assays showed that all five subunits are required to obtain the maximum activity of BAM<sup>13–16</sup>. Individual structures of BamA<sup>17–20</sup>, BamB<sup>21–25</sup>, BamC<sup>22,26</sup>, BamD<sup>22,27,28</sup> and BamE<sup>22,29,30</sup> have previously been reported, as have complex structures of BamD with the N-terminal domain of BamC<sup>31</sup>, and BamB with POTRA 3 and 4 of BamA<sup>32</sup>. Nonetheless, the precise mechanism of OMP insertion by the BAM complex is largely hampered by the lack of a complete structure of the BAM complex<sup>11,32</sup>. Furthermore it is unknown how BAM manages to insert OMPs into the OM without the use of ATP, proton motive forces or redox potentials<sup>33,34</sup>.

Here we report two novel crystal structures of the *E. coli* BAM complex: BamABCDE and BamACDE. The complexes reveal a unique ring architecture that adopts two distinct conformations: an inward-open and a novel lateral-open. Furthermore, comparison of the two complexes reveals that the periplasmic units are rotated with respect to the barrel, which appears to be linked to important conformational changes in the  $\beta$ -strands  $\beta$ 1C– $\beta$ 6C of the barrel. Taken together, this suggests a new insertion mechanism in which rotation of the BAM periplasmic ring promotes insertion of OMPs into the OM. To our knowledge, this is the first reported crystal structure of an intramembrane barrel with a lateral-open conformation.

## Unique architecture of two *E. coli* BAM complexes

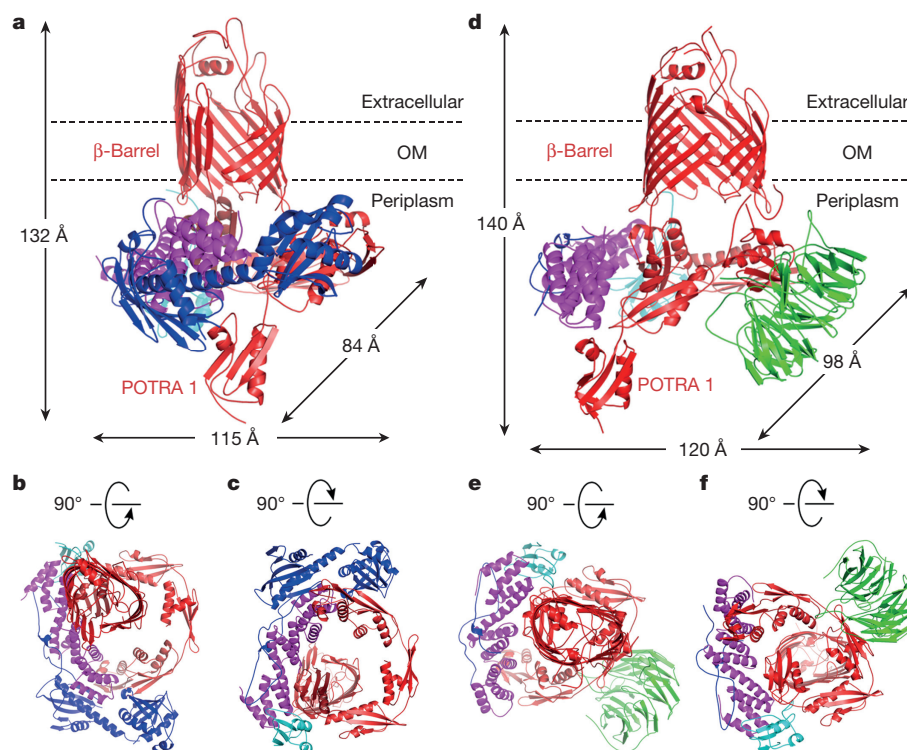
X-ray diffraction data of selenomethionine-labelled crystals were collected to 3.9 Å resolution, and the BAM structure was determined by single-wavelength anomalous dispersion (SAD) and manual molecular replacement (Methods and Extended Data Table 1). The first structure contained four proteins: BamA, BamC, BamD and BamE (Fig. 1a–c), with the electron density and crystal packing indicating that the BamB is absent in the complex. This was confirmed by SDS-PAGE analysis of the crystals (Extended Data Fig. 1 and Supplementary Fig. 1). In this model, BamA, BamC, BamD and BamE contain residues Glu22–Ile806, Cys25–Lys344, Glu26–Ser243 and Cys20–Glu110, respectively. The machinery is approximately 115 Å in length, 84 Å in width and 132 Å in height (Fig. 1a).

The architecture of BamACDE resembles a top hat with an opening in the crown. This crown is formed by the BamA  $\beta$ -barrel with the encircling POTRA domains and associated proteins forming the brim. The C-terminal  $\beta$ -barrel of BamA projects out of the complex and is fully immersed in the OM, while the five POTRA domains of BamA and the BamD form a ring in the periplasm (Fig. 1a–c). The other subunits of the complex surround this central BamAD core. The coiled N-terminal loop of BamC is bound to BamD, as is its

<sup>1</sup>Biomedical Research Centre, Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK. <sup>2</sup>Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, UK. <sup>3</sup>Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. <sup>4</sup>Jiangsu Province Key Laboratory of Anesthesiology, Xuzhou Medical College, Xuzhou 221004, China. <sup>5</sup>Key Laboratory of Bio-resources and Eco-environment, Ministry of Education, Sichuan Key Laboratory of Molecular Biology and Biotechnology, College of Life Sciences, Sichuan University, Chengdu 610064, China. <sup>6</sup>Laboratory of Department of Surgery, the First Affiliated Hospital, Sun Yat-sen University, 58 Zhongshan Road II, Guangzhou, Guangdong 510080, China.

\*These authors contributed equally to this work.





**Figure 1 | Structure of two complexes of *E. coli*  $\beta$ -barrel assembly machinery.** Two structures of *E. coli* BAM: BamABCDE and BamACDE. The BamA (red) C-terminal barrel is embedded in the OM, while the N-terminal domain of BamA is in the periplasm, forming a novel circular structure with lipoproteins BamB (green), BamC (blue), BamD (magenta) and BamE (cyan). **a–c**, Cartoon representation of the structure of the BamACDE complex, viewed for the membrane plane (**a**), extracellular side (**b**) and periplasm (**c**). BamD interacts with POTRA 1, 2 and 5 to form a ring structure in the periplasm, while BamC binds to both BamD and POTRA 1 and 2 of BamA. BamE forms contacts with both BamA and BamD. The dimensions of BamACDE were measured at the widest points of the outer surfaces of the complex. **d–f**, Cartoon representation of BamABCDE structure, viewed from the membrane plane (**d**), extracellular side (**e**) and periplasm (**f**). BamB interacts with POTRA 2 and 3, while only N-terminal loop of BamC forms contacts with BamD. The dimensions of BamABCDE were measured as in **a**.

N-terminal globular domain, which also interacts with POTRA 1 of BamA. The C-terminal globular domain of BamC interacts with BamD and POTRA 2. Notably, for one of the two BamACDE complexes in the asymmetric unit cell, no electron density was observed for the N-terminal and C-terminal globular domains of BamC, which may indicate inherent flexibility (Extended Data Fig. 2). This flexibility was also observed in the molecular dynamics simulations. Finally, BamE is found at the opposite end of the complex, coupling the C-terminal domain of BamD to POTRA 4 and 5 of BamA, adjacent to the barrel (Fig. 1b, c).

To obtain a structure of the complex with all five subunits, we increased the expression level of BamB (Methods). The structure of BamABCDE was determined by co-crystallization with sodium iodine and SAD and manual molecular replacement techniques to a resolution of 2.9 Å (Methods and Extended Data Table 1). The structure we describe below is based on this BamABCDE complex unless otherwise mentioned. The  $\beta$ -strands of the C-terminal barrel of BamA are named  $\beta$ 1C– $\beta$ 16C for consistency with previous reports. The top hat architecture of BamABCDE is similar to that of BamACDE with dimensions of around 120 Å in length, 98 Å in width and 140 Å in height, with the periplasmic ring structure retained (Fig. 1d–f). In the BamABCDE structure, the opening in the crown of the top hat is now closed. This model of the BAM complex contains BamB (residues Lys31–Thr391), which is shown to bind to POTRA 2 and 3. Although SDS–PAGE analysis of the crystals showed BamC is intact in the BamABCDE crystals (Extended Data Fig. 1), electron density is only visible for the N-terminal loop (residues Val35 to Pro88), bound to BamD. This indicates that the rest of BamC is highly flexible. Molecular dynamics simulations of the BamABCDE and BamACDE complexes suggest that both complexes are otherwise stable and the periplasmic ring structure remains intact during the simulations (Extended Data Fig. 3, Supplementary Video 1 and Supplementary Figs 2–4).

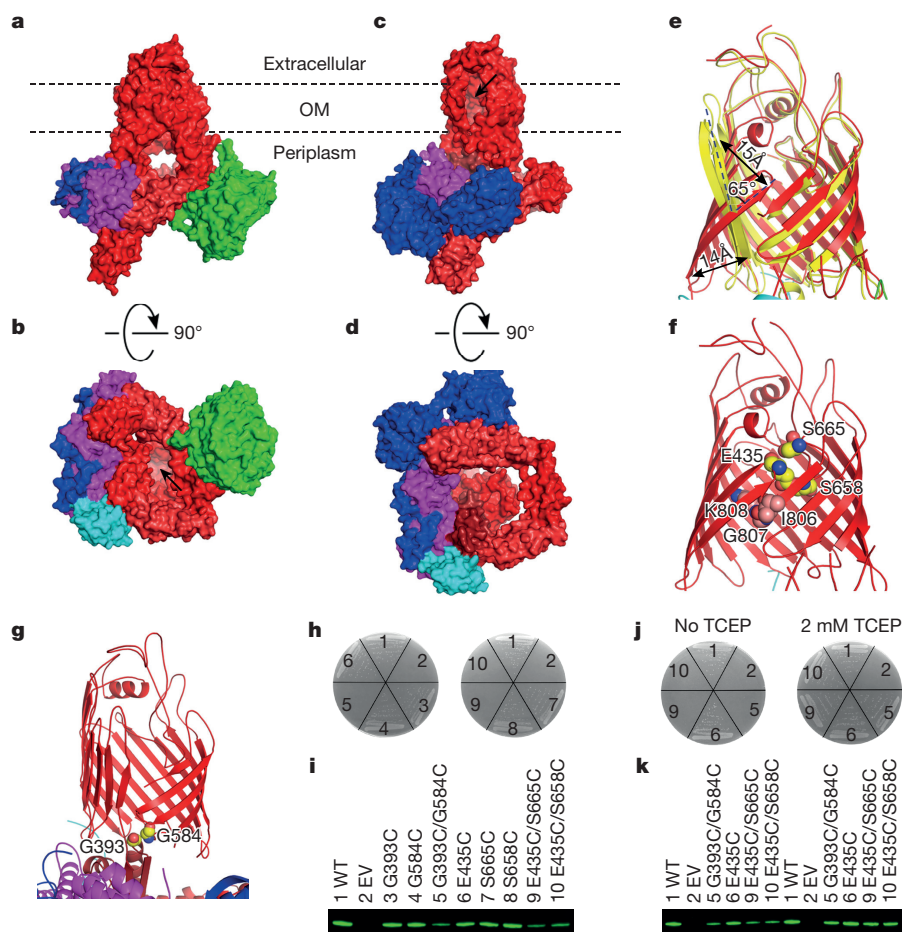
### Inward- and lateral-open conformations

In the structure of the BamABCDE complex, the extracellular loops (L-1 to L-8) cap the pore of BamA to completely close it to the extracellular side, while the periplasmic mouth is fully open to the periplasm (Fig. 2a, b). This conformation is similar to all reported barrel

structures of BamA, however, the POTRA domains are significantly different (Extended Data Fig. 4). The POTRA domains of BamABCDE appear locked through their interactions with BamD, which together form a ring apparatus that may feed the unfolded OMP into the assembly machinery. It is worth noting that the  $\beta$ 16C of both *Neisseria gonorrhoeae* BamA and the BamABCDE complex coils towards the inside of the barrel lumen, creating a gap between  $\beta$ 1C and  $\beta$ 15C of the barrels (Figs. 1d, 2f and Extended Data Fig. 4), which may provide a path for insertion of the OMPs.

By contrast, in the structure of BamACDE, extracellular loops L-1, L-2 and L-3 are displaced from the pore, opening the barrel laterally between  $\beta$ 1C and  $\beta$ 16C. This exposes the barrel lumen to both the extracellular leaflet of the OM and the outside of the cell (Fig. 2c, d). Conversely, on the periplasmic side, POTRA 5 and turn T-1 to T-4 completely plug the barrel (Fig. 2d and Extended Data Fig. 5). The barrel of BamA in the BamACDE structure is therefore in a lateral-open conformation. The first six  $\beta$ -strands,  $\beta$ 1C to  $\beta$ 6C, perform a scissor-like movement to rotate away from the pore to a maximum angle of around 65° and distance of about 15 Å (Fig. 2e and Extended Data Fig. 5). The other strands of the barrel remain unchanged. These conformational changes open the barrel laterally to the OM and the extracellular side, and, in conjunction with POTRA 5, close the periplasmic mouth (Fig. 2c, d and Extended Data Fig. 5). Such a mechanism of conformational changes between inward- and outward-open conformations to transport small molecular substrates is common for  $\alpha$ -helical inner membrane protein transporters<sup>35</sup>. However, to our knowledge, this is the first crystal structure report so far of a  $\beta$ -barrel that may alternate between both inward- and lateral-open conformations. The novel architecture of the lateral-open conformation is likely to facilitate the insertion of  $\beta$ -strands of the OMP into the OM, while permitting the interlinking extracellular loops to extend out of the cell upon insertion.

It was suggested that lateral separation between  $\beta$ 1C and  $\beta$ 16C is required for normal BamA function by disulfide bond cross-linking<sup>36</sup>. To test the two solved conformations, *in vivo* cross-links were designed to interlock BamA in one of the two conformational states. Two double cysteine mutations, Glu435Cys/Ser658Cys and Glu435Cys/Ser665Cys, were created to capture BamA in the inward-open conformation



**Figure 2 | Inward- and lateral-open conformations of BAM.** BamA–BamE are in the same colours as in Fig. 1. The functional assays and the western blots were repeated at least three times. **a**, Membrane view of the molecular surface of BamABCDE. The pore of BamA is completely sealed at the extracellular side by the extracellular loops. **b**, Periplasmic view of BamABCDE. The barrel is open to the periplasm (indicated by the arrow). **c**, Membrane view of the molecular surface of BamACDE. The barrel is open laterally to the OM and the extracellular side (indicated by the arrow). **d**, Periplasmic view of BamACDE surface structure. The barrel is completely closed to the periplasm. **e**, The important conformational changes of the BamA barrel domain between the inward-open (red) and the lateral-open (yellow) conformations. The barrel strands  $\beta$ 1C– $\beta$ 6C of BamA have been rotated about  $65^\circ$  with the distance around  $15\text{ \AA}$  to open the barrel laterally

(Fig. 2f), and one double mutation, Gly393Cys/Gly584Cys, was produced to restrain BamA in the lateral-open conformation (Fig. 2g). The single cysteine mutations do not affect cell growth, while the double cysteine mutations Gly393Cys/Gly584Cys, Glu435Cys/Ser665Cys and Glu435Cys/Ser658Cys are all lethal (Fig. 2h, i). In addition, the double cysteine mutants are folded into the OM, and can be rescued by the addition of 2 mM reducing reagent Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (Fig. 2j, k, Extended Data Fig. 5 and Supplementary Fig. 5), which breaks the disulfide bonds and therefore unlocks the structure, providing strong evidence that the barrel can exist in the two resolved conformations in the bacterial OM.

### Essential interactions between BamA and BamD

Previous mutagenesis analysis has suggested that only POTRA 5 of BamA associates with BamD<sup>37</sup>, however, before this study no structures of this complex have been solved. In our structures, 12 residues of BamD interact with 17 residues of POTRA 5 (Fig. 3a, Extended Data Fig. 6 and Supplementary Table 1). In addition to contacts with POTRA 5, our structures reveal that BamD also interacts with Val480

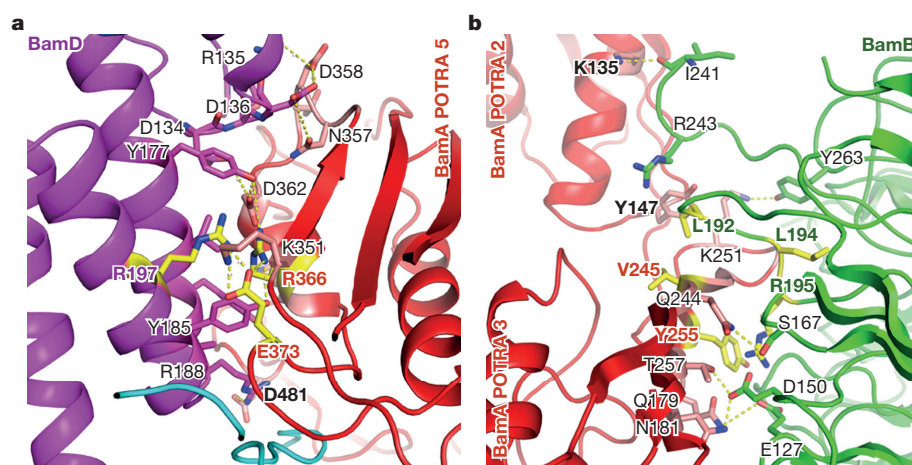
from the inward-open state. **f**, The double mutation Glu435Cys/Ser665Cys or Glu435Cys/Ser658Cys is expected to lock the barrel in the inward-open conformation. Residues Ile806–Lys808 of the  $\beta$ 16C of BamA coils towards the inside of the barrel lumen. **g**, The Gly393Cys/Gly584Cys mutant is expected to lock the barrel in the lateral-open conformation. **h**, The functional assays of the mutants. The single residue mutations do not affect the *E. coli* cell growth, but the double cysteine mutations kill the bacteria. Numbers 1–10 are as shown in **i**. **i**, The protein expression levels of BamA mutants in the OM were checked by western blotting. EV, empty vector without BamA; WT, wild type. **j**, The reducing reagent TCEP could rescue the double cysteine mutations at 2 mM. **k**, The protein expression levels of BamA double cysteine mutants in absence and in presence of TCEP were checked by western blotting.

and Asp481 of periplasmic turn T-2 of the BamA barrel and also forms contacts with POTRA 1 and 2. These interactions complete the ring structure (Extended Data Fig. 7 and Supplementary Table 1).

Molecular dynamics simulations of only the core BamAD periplasmic ring from both structures retain the cyclic complex. Removal of BamD markedly increases the dynamics of POTRA 1 and 2 in the BamACDE conformation (Extended Data Fig. 3 and Supplementary Video 1). In this instance, the POTRA domains rotate in an anticlockwise direction towards the OM. This rotation also causes POTRA 3 to separate from the T-5 and T-6 periplasmic turns of the BamA barrel (Extended Data Fig. 3 and Supplementary Video 1). By contrast, simulations of only BamA from the BamABCDE structure, results in POTRA 2 coupling to POTRA 5, thereby stabilizing POTRA 1 and 2. However, to achieve this configuration, POTRA 3 and 4 separate from the barrel, with a degree of deformation to POTRA 3, suggesting that this is a strained conformation.

To test whether the BamA and BamD interactions are required for BAM function, BamA POTRA 5 mutants Glu373Lys and Arg366Glu were generated. Functional assays showed that Arg366Glu severely





**Figure 3 | BamA interacts with BamD and BamB in BamABCDE complex.** BamD contacts POTRA domains 1, 2 and 5 to form a ring structure. **a**, BamA POTRA 5 interacts with the C-terminal domain of BamD. BamA residues Arg366 and Glu373 and BamD residue Arg197 are important for the two protein interactions, and their carbon atoms are coloured in yellow. **b**, BamA and BamB interaction. Both POTRA 2 and 3 involve in BamB interaction. BamA residues Val245, Tyr255 and BamB residues Leu192, Leu194 and Arg195 have important roles in BamA and BamB interactions.

impairs cell growth, whereas Glu373Lys is lethal to the *E. coli* cells (Extended Data Fig. 7). In the structures, BamA Glu373 and BamD Arg197 form a salt-bridge (Fig. 3a). An Arg197Leu BamD mutation was able to rescue BamA Glu373Lys<sup>37</sup>.

### BamB regulates BamA conformation

The most apparent differences between the two solved structures are the presence of BamB in the BamABCDE complex, while BamC is more clearly resolved in the BamACDE complex (Fig. 2a, c). In addition, the POTRA domains of BamA are found in two distinct conformations, with a larger separation observed between POTRA 3 and 5 in the BamACDE complex. Speculatively, this could be due to the absence of BamB, which binds to POTRA 2 and 3 in the BamABCDE complex and is known to have a regulatory role<sup>3,22,23</sup>. The overall interface between BamA and BamB is around 1,080 Å<sup>2</sup>, and is comprised of the three β-strands of POTRA 3 and a loop consisting of residues Thr245–Lys251 that anchors to the centre of the BamB β-propeller (Fig. 3b, Extended Data Fig. 7 and Supplementary Table 2). The BamB loops at the BamA binding side adopt conformational changes to bind to POTRA 3 (refs 21, 22; Extended Data Fig. 7), consistent with the BamB structure in complex with POTRA 3 and 4 (ref. 32). BamB also interacts with residues Lys135 and Tyr147 of POTRA 2 (Fig. 3b). As a result, the binding of BamB appears to induce local conformational changes in POTRA 2 and 3 with a root mean square deviation (r.m.s.d.) of 3.57 Å over 159 Cα atoms (Extended Data Fig. 7).

Both periplasmic turns T-5 and T-6 of the BamA barrel are more ordered in the BamABCDE structure and interact with POTRA 3 (Extended Data Fig. 6). In the BamACDE complex, POTRA 3 separates from the periplasmic turns (Fig. 2c), indicating that BamB may have a role in controlling the structural rearrangements of the barrel through POTRA 3. It is worth noting that POTRA 5 also has extensive contacts with the periplasmic turns T-1, T-2 and T-3 of the barrel domain in the BamABCDE structure (Extended Data Fig. 6), and we speculate that the conformational changes of POTRA domains may play a role in controlling the conformations of the barrel.

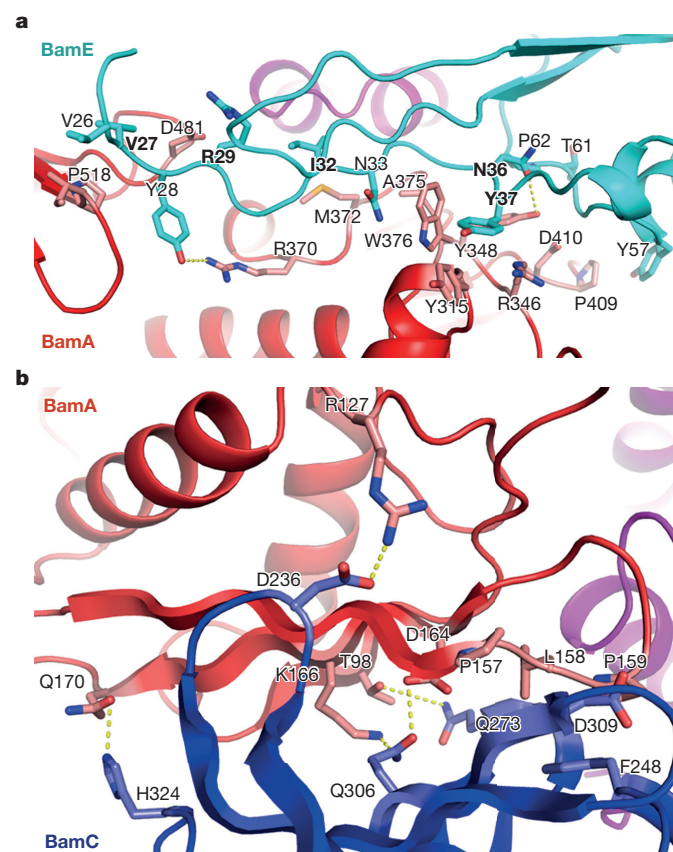
In the molecular dynamics simulations, the tight coupling between POTRA 3 and BamB was retained. However, in the absence of BamB, the simulations reveal greater dynamics of POTRA 3 and 4, with both domains moving away from the barrel and the membrane (Supplementary Video 1). This suggests that BamB is important for coupling of POTRA 3 at the appropriate height with respect to the barrel and OM.

### BamE and BamC interactions with BamA and BamD

Previous studies suggested that BamE only binds to BamD directly<sup>38,39</sup>. Surprisingly, both BamABCDE and BamACDE structures show that BamE is not only positioned between BamA and BamD, but also forms contacts with BamC (Fig. 4a, Extended Data Figs 6 and 8 and

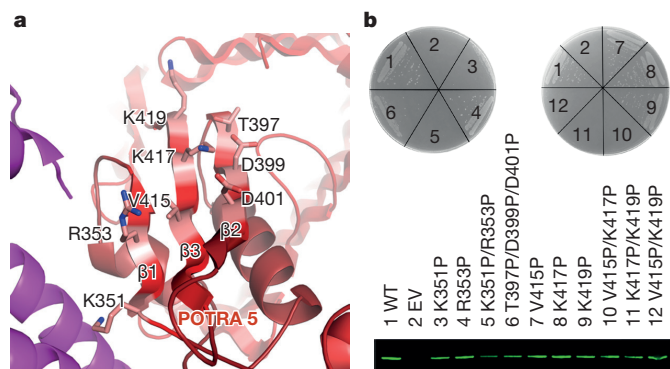
Supplementary Tables 3 and 4). Notably, the BamE residues Pro67 and Phe68 also interact with BamC residues Met56 and Ile57 (Extended Data Fig. 8), suggesting that BamC, BamD and BamE form a network to regulate the conformations of BamA.

BamE residues Ile32, Gln34, Leu63 and Arg78 interact with both BamA and BamD. Mutations to any of these residues caused defects to the OM<sup>29</sup>. Additionally, single BamE mutations Asn36 and Tyr37 at the BamE and BamA interface or residues Met64, Asp66, Phe74, Val76 and Gln88 at the BamE and BamD interface caused defects of



**Figure 4 | BamE and BamC interact with BamA and BamD.** **a**, The interface between BamE and BamA in the BamABCDE complex. BamE forms contacts with POTRA 5 residues, BamA periplasmic turns T-2 and T-3, and POTRA 4 in the BamACDE complex (Extended Data Fig. 6). **b**, The C-terminal globular domain of BamC interacts with BamA POTRA 2 at the β-sheets in BamACDE. Residues in the two β-sheets that are involved in the BamC and BamA interactions are shown.





**Figure 5 | The function of BamA POTRA 5.** **a**, The three  $\beta$ -strands of BamA POTRA 5. The residues selected for functional assays are shown: residues Lys351 and Arg353 on  $\beta$ 1, Thr397, Asp399 and Asp401 on  $\beta$ 2 and Val415, Lys417 and Lys419 on  $\beta$ 3. **b**, The  $\beta$ -strands of POTRA 5 are crucial for bacterial survival. The functional assays and the western blots were repeated at least three times. The single mutant Lys351Pro, the double mutant Lys351Pro/Arg353Pro at  $\beta$ 1, the double mutants Val415Pro/Lys417Pro/Lys419Pro/Lys419Pro and Val415Pro/Lys419Pro at  $\beta$ 3 kill the bacteria, while the triple mutant Thr397Pro/Asp399Pro/Asp401Pro at  $\beta$ 2 impairs cell growth. The protein expression of the BamA wild-type and mutants. Numbers 1–12 are as shown below.

the OM<sup>29</sup>. These data suggest that BamE has an important role in OMP assembly.

The whole BamC structure is revealed in the BamACDE complex. BamC forms extensive contacts with BamD, with the average interface of 2,686 Å<sup>2</sup> (Fig. 4b, Extended Data Fig. 8 and Supplementary Table 5). The N-terminal loop of BamC, up to residue Gly94, is largely unstructured, coiling round BamD and forming a cluster of contacts with BamE. The N-terminal globular domain of BamC interacts with the N-terminal domain of BamD and POTRA 1 of BamA (Extended Data Fig. 8). The C-terminal globular domain interacts principally with POTRA 2, via their  $\beta$ -sheets (Fig. 4b).

The C-terminal globular domain of BamC binds to POTRA 2 in one of the BamACDE complexes. This probably enhances the periplasmic ring structure formed by BamA and BamD, and has a role in the conformational changes of BAM. Comparison of the two BamACDE complexes in the asymmetric unit reveals differences in the barrel domain, with an r.m.s.d. of 0.91 Å over 378 C $\alpha$  atoms, while the periplasmic ring is somewhat rotated, with respect to the barrel (Extended Data Fig. 2). SDS-PAGE analysis confirmed that both BamACDE and BamABCDE crystals contain full-length BamC (Extended Data Fig. 1), suggesting that the two globular domains of BamC are dynamic. Molecular dynamics simulations, with the addition of the BamC globular domains to this structure, also show that these domains are not tightly coupled to the complex. Indeed, in the simulations of both complexes, BamC shows the least stability of all five subunits. The total interface between BamC with BamA is around 794 Å<sup>2</sup> in the BamACDE structure (Fig. 4b, Extended Data Fig. 8 and Supplementary Table 6). The molecular dynamics simulations of BamACDE complex without BamC suggest that the POTRA 1 moves towards to the membrane, while POTRA 3 moves towards the barrel and engages with the periplasmic turns T-5 and T-6 (Extended Data Fig. 3). In the simulations of the inward-open BamABCDE complex the absence of BamC globular domains has limited effects on the overall structure, as BamD is more tightly coupled to the POTRA domains. Taken together, all four lipoproteins BamB, C, D and E have direct contacts with BamA POTRA domains, which may be important in terms of conformation and functional regulation. Analysis of the BAM subunits reveals that conserved residues are mapped to those regions involved in inter-subunit protein–protein interactions (Extended Data Fig. 9 and Supplementary Fig. 6).

All the POTRA domains of BamA have the  $\beta\alpha\alpha\beta\beta$  fold<sup>17</sup>. An NMR study suggested that the  $\beta$ -sheet of the POTRA domains may bind

substrate in a non-specific manner<sup>40</sup>. Our structural studies show that both BamB and the C-terminal domain of BamC bind to the POTRA domains 2 and 3 through their  $\beta$ -sheets. The three  $\beta$ -strands of POTRA 5 adopt an important conformational change, from 126° between the  $\beta$ 1C and the POTRA 5  $\beta$ -sheet ( $\beta$ 3) in the inward-open state to 165° in the lateral-open state, aligning the  $\beta$ -sheet of POTRA 5 with  $\beta$ 1C. Furthermore, almost all BamA of Gram-negative bacteria, SAM50 of mitochondria and OEP80 of chloroplasts have the last POTRA domain, indicating that the POTRA 5 of BamA may play an important role in the insertion of OMPs. To test this possibility, single proline substitutions (Lys351Pro, Arg353Pro, Thr397Pro, Asp399Pro, Asp401Pro, Val415Pro, Lys417Pro, Lys419Pro), double proline substitutions (Lys351Pro/Arg353Pro, Thr397Pro/Asp399Pro, Asp399Pro/Asp401Pro, Thr397Pro/Asp401Pro, Val415Pro/Lys417Pro, Lys417Pro/Lys419Pro, Val415Pro/Lys419Pro) and a triple proline substitution (Thr397Pro/Asp399Pro/Asp401Pro) were generated in  $\beta$ 1,  $\beta$ 2 and  $\beta$ 3 of BamA POTRA 5. Functional assays showed that all single mutants (except Lys351Pro) and double mutants at  $\beta$ 2 (Thr397Pro/Asp399Pro, Asp399Pro/Asp401Pro, Thr397Pro/Asp401Pro) did not affect *E. coli* cell growth, but single substitution Lys351Pro at  $\beta$ 1, double proline substitutions at  $\beta$ 1 (Lys351Pro/Arg353Pro) and  $\beta$ 3 (Val415Pro/Lys417Pro, or Lys417Pro/Lys419Pro, or Val415Pro/Lys419Pro) are lethal, and the triple mutation at  $\beta$ 2 (Thr397Pro/Asp399Pro/Asp401Pro) impaired the cell growth (Fig. 5a, b). This strongly suggests that the  $\beta$ -sheet, especially  $\beta$ -strands 1 and 3 of POTRA 5, may have a critical role in OMPs insertion, possibly by  $\beta$ -augmentation of the unfolded OMPs.

## Mechanism and conclusion

In Gram-negative bacteria, outer membrane barrel proteins are inserted and assembled into the OM by the BAM complex. Our studies have revealed the three-dimensional architecture of the entire *E. coli* BAM complex, trapped in two distinct conformational states. The structures suggest that a rotation of the periplasmic ring (Extended Data Figs 2, 5 and Supplementary Videos 1, 2) and conformational changes of the POTRA domains and BamB–BamE (Extended Data Fig. 10) induces the considerable conformational changes to the barrel of BamA required for BAM-induced OMP insertion (Fig. 2e). Considering all four lipoproteins subunits, BamB–BamE, directly interact with POTRA domains, the ring architecture of the *E. coli* BAM complex may be an efficient way to coordinate all BAM subunits and thereby promote OMPs insertion into the OM (Extended Data Fig. 10 and Supplementary Video 2).

To accomplish insertion, the OM periplasmic lipid head groups must be circumnavigated by the unfolded or partially folded OMPs<sup>41</sup>. Several mechanisms for OMP insertion have previously been described<sup>18,36</sup>, with the 'BamA-assisted model' and the 'budding model' currently the two most favoured<sup>36</sup>.

Our structures reveal a 30° rotation of the periplasmic ring complex, which interacts directly with the lipid headgroups of the OM (Extended Data Figs 5, 10). This rotation is probably coupled to the 65° tilting of strands  $\beta$ 1C– $\beta$ 6C of the BamA barrel and the partial separation of the lateral gate, formed by  $\beta$ 1C and  $\beta$ 16C (Fig. 2c, e). This exposes the barrel lumen to the core of the OM, while also inducing a degree of membrane instability to facilitate OMP insertion. The BamA homologue, SAM50, in mitochondria will probably use a similar scissor-like movement of the barrel strands to promote OMPs insertion into the mitochondrial OM; however, this is performed in absence of the periplasmic ring.

In summary, our structural, functional and molecular dynamics simulations have revealed that the BAM complex has a unique ring architecture and is able to adopt both inward-open and lateral-open states. We propose that these structures represent the resting (BamABCDE) and post-insertion (BamACDE) states of the complex. These findings shed an important light on how the BAM subunits work together to insert unfolded OMPs into the OM without using ATP, and sets up

an important platform for further studies of OM biogenesis and the potential development of novel therapies, for example by inhibiting complex formation.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 3 August 2015; accepted 5 February 2016.**

**Published online 22 February 2016.**

- Walther, D. M., Rapaport, D. & Tommassen, J. Biogenesis of  $\beta$ -barrel membrane proteins in bacteria and eukaryotes: evolutionary conservation and divergence. *Cell. Mol. Life Sci.* **66**, 2789–2804 (2009).
- Tommassen, J. Assembly of outer-membrane proteins in bacteria and mitochondria. *Microbiology* **156**, 2587–2596 (2010).
- Wu, T. et al. Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. *Cell* **121**, 235–245 (2005).
- Sasaki, K. et al. VDAC: Old protein with new roles in diabetes. *Am. J. Physiol. Cell Physiol.* **303**, C1055–C1060 (2012).
- Bender, A. et al. TOM40 mediates mitochondrial dysfunction induced by  $\alpha$ -synuclein accumulation in Parkinson's disease. *PLoS ONE* **8**, e62277 (2013).
- Knowles, T. J., Scott-Tucker, A., Overduin, M. & Henderson, I. R. Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nature Rev. Microbiol.* **7**, 206–214 (2009).
- Hagan, C. L., Silhavy, T. J. & Kahne, D.  $\beta$ -barrel membrane protein assembly by the Bam complex. *Annu. Rev. Biochem.* **80**, 189–210 (2011).
- Ricci, D. P. & Silhavy, T. J. The Bam machine: a molecular cooper. *Biochim. Biophys. Acta* **1818**, 1067–1084 (2012).
- Rigel, N. W. & Silhavy, T. J. Making a  $\beta$ -barrel: assembly of outer membrane proteins in Gram-negative bacteria. *Curr. Opin. Microbiol.* **15**, 189–193 (2012).
- Webb, C. T., Heinz, E. & Lithgow, T. Evolution of the  $\beta$ -barrel assembly machinery. *Trends Microbiol.* **20**, 612–620 (2012).
- Noinaj, N., Rollauer, S. E. & Buchanan, S. K. The  $\beta$ -barrel membrane protein insertase machinery from Gram-negative bacteria. *Curr. Opin. Struct. Biol.* **31**, 35–42 (2015).
- Misra, R., Stikeleather, R. & Gabriele, R. *In vivo* roles of BamA, BamB and BamD in the biogenesis of BamA, a core protein of the  $\beta$ -barrel assembly machine of *Escherichia coli*. *J. Mol. Biol.* **427**, 1061–1074 (2015).
- McMorran, L. M., Brockwell, D. J. & Radford, S. E. Mechanistic studies of the biogenesis and folding of outer membrane proteins *in vitro* and *in vivo*: what have we learned to date? *Arch. Biochem. Biophys.* **564**, 265–280 (2014).
- Hagan, C. L., Kim, S. & Kahne, D. Reconstitution of outer membrane protein assembly from purified components. *Science* **328**, 890–892 (2010).
- Hagan, C. L., Westwood, D. B. & Kahne, D. Bam lipoproteins assemble BamA *in vitro*. *Biochemistry* **52**, 6108–6113 (2013).
- Roman-Hernandez, G., Peterson, J. H. & Bernstein, H. D. Reconstitution of bacterial autotransporter assembly using purified components. *Elife* **3**, e04234 (2014).
- Kim, S. et al. Structure and function of an essential component of the outer membrane protein assembly machine. *Science* **317**, 961–964 (2007).
- Noinaj, N. et al. Structural insight into the biogenesis of  $\beta$ -barrel membrane proteins. *Nature* **501**, 385–390 (2013).
- Ni, D. et al. Structural and functional analysis of the  $\beta$ -barrel domain of BamA from *Escherichia coli*. *FASEB J.* **28**, 2677–2685 (2014).
- Albrecht, R. et al. Structure of BamA, an essential factor in outer membrane protein biogenesis. *Acta Crystallogr. D* **70**, 1779–1789 (2014).
- Noinaj, N., Fairman, J. W. & Buchanan, S. K. The crystal structure of BamB suggests interactions with BamA and its role within the BAM complex. *J. Mol. Biol.* **407**, 248–260 (2011).
- Albrecht, R. & Zeth, K. Structural basis of outer membrane protein biogenesis in bacteria. *J. Biol. Chem.* **286**, 27792–27803 (2011).
- Dong, C., Yang, X., Hou, H. F., Shen, Y. Q. & Dong, Y. H. Structure of *Escherichia coli* BamB and its interaction with POTRA of BamA. *Acta Crystallogr. D* **68**, 1134–1139 (2012).
- Kim, K. H. & Paetzel, M. Crystal structure of *Escherichia coli* BamB, a lipoprotein component of the  $\beta$ -barrel assembly machinery complex. *J. Mol. Biol.* **406**, 667–678 (2011).
- Heuck, A., Schleiffer, A. & Clausen, T. Augmenting  $\beta$ -augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins. *J. Mol. Biol.* **406**, 659–666 (2011).
- Warner, L. R. et al. Structure of the BamC two-domain protein obtained by Rosetta with a limited NMR data set. *J. Mol. Biol.* **411**, 83–95 (2011).
- Sandoval, C. M., Baker, S. L., Jansen, K., Metzner, S. I. & Sousa, M. C. Crystal structure of BamD: an essential component of the  $\beta$ -barrel assembly machinery of gram-negative bacteria. *J. Mol. Biol.* **409**, 348–357 (2011).
- Dong, C., Hou, H. F., Yang, X., Shen, Y. Q. & Dong, Y. H. Structure of *Escherichia coli* BamD and its functional implications in outer membrane protein assembly. *Acta Crystallogr. D* **68**, 95–101 (2012).
- Knowles, T. J. et al. Structure and function of BamE within the outer membrane and the  $\beta$ -barrel assembly machine. *EMBO Rep.* **12**, 123–128 (2011).
- Kim, K. H. et al. Structural characterization of *Escherichia coli* BamE, a lipoprotein component of the  $\beta$ -barrel assembly machinery complex. *Biochemistry* **50**, 1081–1090 (2011).
- Kim, K. H., Aulakh, S. & Paetzel, M. Crystal structure of  $\beta$ -barrel assembly machinery BamCD protein complex. *J. Biol. Chem.* **286**, 39116–39121 (2011).
- Jansen, K. B., Baker, S. L. & Sousa, M. C. Crystal structure of BamB bound to a periplasmic domain fragment of BamA, the central component of the  $\beta$ -barrel assembly machine. *J. Biol. Chem.* **290**, 2126–2136 (2015).
- Hagan, C. L. & Kahne, D. The reconstituted *Escherichia coli* Bam complex catalyzes multiple rounds of  $\beta$ -barrel assembly. *Biochemistry* **50**, 7444–7446 (2011).
- Rigel, N. W., Ricci, D. P. & Silhavy, T. J. Conformation-specific labeling of BamA and suppressor analysis suggest a cyclic mechanism for  $\beta$ -barrel assembly in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **110**, 5151–5156 (2013).
- Noinaj, N. & Buchanan, S. K. Structural insights into the transport of small molecules across membranes. *Curr. Opin. Struct. Biol.* **27**, 8–15 (2014).
- Noinaj, N., Kuszak, A. J., Balusek, C., Gumbart, J. C. & Buchanan, S. K. Lateral opening and exit pore formation are required for BamA function. *Structure* **22**, 1055–1062 (2014).
- Ricci, D. P., Hagan, C. L., Kahne, D. & Silhavy, T. J. Activation of the *Escherichia coli*  $\beta$ -barrel assembly machine (Bam) is required for essential components to interact properly with substrate. *Proc. Natl Acad. Sci. USA* **109**, 3487–3491 (2012).
- Rigel, N. W., Schwalm, J., Ricci, D. P. & Silhavy, T. J. BamE modulates the *Escherichia coli*  $\beta$ -barrel assembly machine component BamA. *J. Bacteriol.* **194**, 1002–1008 (2012).
- Tellez, R., Jr & Misra, R. Substitutions in the BamA  $\beta$ -barrel domain overcome the conditional lethal phenotype of a  $\Delta$ bamB  $\Delta$ bamE strain of *Escherichia coli*. *J. Bacteriol.* **194**, 317–324 (2012).
- Bennion, D., Charlson, E. S., Coon, E. & Misra, R. Dissection of  $\beta$ -barrel outer membrane protein assembly pathways through characterizing BamA POTRA 1 mutants of *Escherichia coli*. *Mol. Microbiol.* **77**, 1153–1171 (2010).
- Gessmann, D. et al. Outer membrane  $\beta$ -barrel protein folding is physically controlled by periplasmic lipid head groups and BamA. *Proc. Natl Acad. Sci. USA* **111**, 5878–5883 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank H. D. Bernstein for providing HDB150 strain and plasmid pJH114 and T. J. Silhavy for providing JCM166 cells. We appreciate the staff at I24, I02 and I03 of Diamond Light Source UK for beamtime (proposal mx9475) and their assistance with data collection. C.D. is a recipient of the Wellcome Trust investigator award (WT106121MA), and is supported by Medical research council (G1100110/1). W.W. acknowledges the support of the Science and Technology Program of Guangzhou, China (201510010040) and China National Natural Science Foundation of Guangdong (2015A030313152).

**Author Contributions** C.D., W.W. and Y.G. conceived and designed the experiments. C.D. and W.W. supervised the project. Y.G., Y.Ze, Z.W. and Y.Zh designed primers and generated the constructs for protein expression and functional assays. Z.W., Y.G., Y.Ze, H.D., H.L. and Z.Z. participated in expression, purification and crystallization of the BAM complex. Y.G. and Y.Ze expressed the BamACDE complex using the plasmids pJH114, and expressed the BamABCDE complex using the plasmid pYG120. Y.G. and Y.Ze purified and crystallized the complexes, optimized crystallization, obtained well diffracted crystals and performed site-directed mutagenesis, functional assays, heat-modifiability assays and western blot. N.G.P., Y.G., C.D. and W.W. undertook data collection and structure determination of the BamACDE and BamABCDE complexes. P.J.S. performed the molecular dynamics simulations. Y.G., C.D., P.J.S. and N.G.P. prepared tables and figures. C.D., Y.G., W.W., P.J.S. and N.G.P. wrote and revised the manuscript.

**Author Information** The atomic coordinates and structure factors of BamABCDE and BamACDE have been deposited in the Protein Data Bank under accession codes 5D00 and 5D0Q. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.D. (c.dong@uea.ac.uk) or W.W. (w166w2000@aliyun.com).



## METHODS

**Cloning, expression and purification of BAM complex.** Expression plasmid pJH114 containing the five *E. coli* *bamABCDE* genes which were under the control of a *trc* promoter, and with an octa-histidine ( $8 \times \text{His}$ ) tag at the C terminus of *bamE* was initially used for overexpression of BamABCDE complex in *E. coli* HDB150 cells<sup>16</sup>. Expression of the native BamABCDE complex was induced with  $100 \mu\text{mol l}^{-1}$  isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG; Formedium) at  $20^\circ\text{C}$  overnight when the absorbance of the cell culture at 600 nm reached 0.5–0.8. The selenomethionine-labelled BAM complexes were expressed in M9 medium supplemented with selenomethionine Medium Nutrient Mix (Molecular Dimensions) and  $100 \text{ mg l}^{-1}$  L-(+)-selenomethionine (Generon) using the similar conditions as the native BamABCDE.

Both native and selenomethionine-labelled BamABCDE complexes were purified using a similar protocol. In brief, the cells were pelleted and resuspended in lysis buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM NaCl,  $10 \mu\text{g ml}^{-1}$  DNase I and  $100 \mu\text{g ml}^{-1}$  lysozyme and lysed by passing through a cell disruptor (Constant Systems) at 206 MPa. The lysate was centrifuged to remove the cell debris and unbroken cells, and the supernatant was ultracentrifuged to pellet the membranes at  $100,000g$  for 1 h. The cell membranes were resuspended in solubilization buffer containing 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10 mM imidazole and 1–2% *n*-dodecyl- $\beta$ -D-maltopyranoside (DDM; all detergents were purchased from Anatrace) and rocked for 1 h at room temperature or overnight at  $4^\circ\text{C}$ . The suspension was ultracentrifuged and the supernatant was applied to a 5-ml pre-equilibrated HisTrap HP column (GE Healthcare). The column was washed with wash buffer containing 20 mM Tris-HCl, pH 8.0, 300 mM NaCl and 35 mM imidazole and eluted with elution buffer containing 300 mM imidazole. The eluent was applied to HiLoad 16/600 Superdex 200 prep grade column (GE healthcare) pre-equilibrated with gel filtration buffer containing 20 mM Tris-HCl, pH 7.8, 300 mM NaCl and detergents. Different detergents were used in protein purification procedures.

The purified BamABCDE complex was analysed by SDS-PAGE (Extended Data Fig. 1 and Supplementary Fig. 1), which indicated that BamB is not enough in the complex, and BamB is absent in the determined structure. We therefore decided to generate a new plasmid to express the BamABCDE complex. Additional copy of the *E. coli* *bamB* gene was introduced into pJH114 (ref. 16) after the  $8 \times \text{His}$  tag to generate a new expression plasmid pYG120 using a modified sequence and ligation-independent cloning (SLIC) method<sup>42</sup>. In brief, vector backbone and *bamB* gene fragments were amplified by PCR using Q5 Hot Start High-Fidelity DNA Polymerase (New England BioLabs), and plasmid pJH114 as template and primers PF\_pJH114\_SLIC (5'-GTTAATCGACC TGCAGGCATGCAAG-3') and PR\_pJH114\_SLIC (5'-CTCTAGAGGATCTTAG TGGTGATGATGGTG-3'), and PF\_EBB\_SLIC (5'-TCATCACCCTAAGATCCT CTAGAGAGGGACCCGATGCAATTGC-3') and PR\_EBB\_SLIC (5'-CTT GCATGCCTGCAGGTCGATTAACTGTAATAGAGTACACGGTTCC-3'), respectively. Gel-extracted fragments were digested by T4 DNA polymerase (Fermentas) at  $22^\circ\text{C}$  for 35 min followed by  $70^\circ\text{C}$  for 10 min, and then placed on ice immediately. The digested fragments were annealed in an annealing buffer (10 mM Tris, pH 8.0, 100 mM NaCl and 1 mM EDTA) by incubating at  $75^\circ\text{C}$  for 10 min and decreasing by  $0.1^\circ\text{C}$  every 8 s to  $20^\circ\text{C}$ . The mixture was transformed into *E. coli* DH5 $\alpha$  for plasmid preparation. The DNA sequences were confirmed by sequencing. For the purification of the BamABCDE complex from the pYG120 construct, the wash buffer, elution buffer and gel filtration buffer were supplemented with different detergent combinations. A second gel filtration was performed to change detergents with gel filtration buffer containing 1 CMC *N*-octyl- $\beta$ -D-glucopyranoside (OG) and 1 CMC *N*-dodecyl-*N,N*-dimethylamine-*N*-oxide (LDAO). For BamABCDE complex purification from construct pJH114, the wash buffer, elution buffer and gel filtration buffer were supplemented with 2 CMC *N*-nonyl- $\beta$ -D-glucoside ( $\beta$ -NG) and 1 CMC tetraethylene glycol mono-octyl ether (C8E4). The peak fraction was pooled and concentrated using Vivaspin 20 centrifugal concentrator (Sartorius, molecular mass cut off: 100 kDa). The selenomethionine-labelled proteins were purified in the same way as the native proteins of BamABCDE complex.

**Crystallization, data collection and structure determination.** The purified proteins were concentrated to  $8\text{--}12 \text{ mg ml}^{-1}$  for crystallization. For NaI co-crystallization, NaCl was replaced by NaI in the gel filtration buffer. All crystallizations were carried out by sitting-drop vapour diffusion method in the MRC 96-well crystallization plates (Molecular Dimensions) at  $22^\circ\text{C}$ . The protein solution was mixed in a 1:1 ratio with the reservoir solution using the Gryphon crystallization robot (Art Robbins Instruments). The best NaI co-crystallized crystals were grown from 150 mM HEPES, pH 7.5, 30% PEG6000 and CYMAL-4 in MemAdvantage (Molecular Dimensions) as additive. The best native crystals were grown from 150 mM HEPES, pH 7.5 and 27.5% PEG6000. The best selenomethionine-labelled crystals were grown from 100 mM Tris, pH 8.0, 200 mM  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ ,

24% PEG1000 MME and OGNG in MemAdvantage as additive. The crystals were harvested, flash-cooled and stored in liquid nitrogen for data collection. The data sets of selenomethionine labelled BAM complex were collected on the I03 beamline at Diamond Light Resources (DLS) at a wavelength of  $0.9795 \text{ \AA}$ . All data were indexed, integrated and scaled using XDS<sup>43</sup>. The crystals belong to space group of  $P4_22_12$ , with the cell dimensions  $a = b = 254.16 \text{ \AA}$ ,  $c = 179.22$ ,  $\alpha = \beta = \gamma = 90^\circ$ . There are two complexes in the asymmetric unit. The structure was determined to  $3.9 \text{ \AA}$  resolution (Extended Data Table 1) using ShelxD<sup>44,45</sup>. Fifty-six selenium sites were found, which gave a figure of merit (FOM) of 0.32. After density modification using DM<sup>46</sup>, the BamACDE complex was clearly visible in the electron density map, but without BamB. Using the individual high-resolution models, the BamACDE complex was built using Coot<sup>47</sup> by skeletonizing the electron density map and docking the BAM subunits in the electron density map with selenomethionine sites used as guides. Rigid body refinement was performed following manual docking. NCS refinement was used along with TLS refinement against groups automatically determined using PHENIX<sup>48</sup>. Restrained refinement was performed with group B-factors alongside reference model secondary structure restraints from higher resolution models. Weights were automatically optimised by PHENIX<sup>48</sup>.

To obtain the BamABCDE complex structure, the new construct was used to produce sufficient BamB to form the BamABCDE complex. The data sets of BamABCDE complex were collected on the I02 beamline at DLS. The crystals belong to space group  $P4_22_12$ , with the cell dimensions  $a = b = 116.69 \text{ \AA}$ ,  $c = 435.19 \text{ \AA}$ ,  $\alpha = \beta = \gamma = 90^\circ$ . There is one complex molecule in the asymmetric unit. Although the crystals diffracted to  $2.90 \text{ \AA}$ , the crystal structure of BamABCDE could not be determined by molecular replacement. BamABCDE complex was crystallized in presence of 0.2 M sodium iodide, and SAD data sets were collected at a wavelength of  $1.8233 \text{ \AA}$ . Four  $360^\circ$  data sets were collected on different regions of the same crystal of NaI co-crystallization then combined. The phases were determined by ShelxD<sup>44,45</sup> at  $4 \text{ \AA}$  resolution. Eleven iodide sites were found, which gave a FOM of 0.28. The phases were extended to  $2.90 \text{ \AA}$  by DM<sup>46</sup>, and the model was built using Coot<sup>47</sup> by skeletonizing the electron density map and docking the individual high-resolution subunits in the electron density map and rigid body fit this model into the higher resolution native data set while retaining and extending the free R set from the iodide data set. The BamABCDE complex was refined using PHENIX<sup>48</sup>. TLS groups were automatically determined using PHENIX<sup>48</sup> and used for refinement along with individual B-factors. Weights were automatically optimised and secondary structure restraints were used.

**Site-directed mutagenesis and functional assays.** An *E. coli* *bamA* expression plasmid was constructed for functional assays using SLIC method as described above. An N-terminal  $10 \times \text{His}$  tag fused with *bamA* starting from residue 22 was amplified by PCR using Q5 Hot Start High-Fidelity DNA Polymerase (New England BioLabs), and plasmid pJH114 as template and primers PF\_bamA\_SLIC (5'-CCATCATCATCATCATCATCATCATGAAG GGTTCGTAGTGAAAGATATTCATTTTCGAAG-3') and PR\_bamA\_SLIC (5'-AGACTCGAGTTACAGGTTTTACCGATGTTAACTGGAAC-3'). Vector backbone was amplified from a modified pRSFDuet-1 vector (Novagen, Merck Millipore) containing an N-terminal pelB signal peptide coding sequence with primers PF\_RSFM\_SLIC (5'-CGGTAAAACCTGGTAACCTCGAGTCT GGTAAAGAAACCGCTGC-3') and PR\_RSFM\_SLIC (5'-ATGATGATGAT GATGATGATGATGGTATGGGCCATCGCCGGCTG-3'). Plasmids were prepared using GeneJET Plasmid Miniprep Kit (Thermo Scientific). Site-directed mutagenesis was performed according to a previously described protocol<sup>49</sup> with slight modification (PCR conditions and the sequences of the primers are available on request). The sequences of the wild type and all mutant constructs of BamA were confirmed by sequencing. *E. coli* JCM166 cells<sup>3</sup> transformed with the wild-type BamA or its mutants were plated on LB agar plates supplemented with  $50 \mu\text{g ml}^{-1}$  kanamycin and  $100 \mu\text{g ml}^{-1}$  carbenicillin in the presence or absence of 0.05% L-(+)-arabinose and grown overnight at  $37^\circ\text{C}$ . Single colonies grown on arabinose-containing plates were inoculated in 10 ml LB medium supplemented with  $50 \mu\text{g ml}^{-1}$  kanamycin,  $100 \mu\text{g ml}^{-1}$  carbenicillin and 0.025% L-(+)-arabinose, and incubated at 200 r.p.m. at  $37^\circ\text{C}$  for 16 h. For plate assays, the cells were pelleted and resuspended in fresh LB medium supplemented with  $50 \mu\text{g ml}^{-1}$  kanamycin and  $100 \mu\text{g ml}^{-1}$  carbenicillin, and diluted to an  $A_{600 \text{ nm}}$  of  $\sim 0.3$  and streaked onto LB agar plates supplemented with  $50 \mu\text{g ml}^{-1}$  kanamycin,  $100 \mu\text{g ml}^{-1}$  carbenicillin in the presence or absence of 0.05% L-(+)-arabinose and cultured at  $37^\circ\text{C}$  for 12–14 h.

**Western blot.** Western blotting was performed to examine protein expression levels of BamA in the membrane. 50 ml of overnight cultures of transformed JCM166 cells with respective wild-type or each mutant of BamA were pelleted. The cells were resuspended in 25 ml 20 mM Tris-HCl, pH 8.0, 150 mM NaCl and sonicated. The cell debris and unbroken cells were removed by centrifugation at  $7,000g$  for 30 min. The supernatant was centrifuged at  $100,000g$  for 60 min and the membrane fraction was collected. The membrane fraction was suspended



in 5 ml buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM NaCl and 1% 3-(*N*, *N*-dimethylmyristylammonio)-propanesulfonate (Sigma) and solubilized for 30 min at room temperature. Samples were mixed with 5 × SDS-PAGE loading buffer, heated for 5 min at 90 °C, cooled for 2 min on ice and centrifuged. Ten microlitres of each sample was loaded onto 4–20% Mini-PROTEAN TGX Gel (Bio-Rad) for SDS-PAGE and then subjected to immunoblot analysis.

The proteins were transferred to PVDF membrane using Trans-Blot Turbo Transfer Starter System (Bio-Rad) according to the manufacturer's instructions. The PVDF membranes were blocked in 10 ml protein-free T20 (TBS) blocking buffer (Fisher) overnight at 4 °C. The membranes were incubated with 10 mL His-Tag monoclonal antibody (diluted, 1:1,000) (Millipore) for 1 h at room temperature followed by washed with PBST four times and incubated with IRDye 800CW goat anti-mouse IgG (diluted, 1:5,000) (LI-COR) for 1 h. The membrane was washed with PBST four times and PBS twice. Images were acquired using LI-COR Odyssey (LI-COR).

**BamA heat-modifiability assays.** The JCM166 cells containing the double cysteine mutants Gly393Cys/Gly584Cys, Glu435Cys/Ser665Cys and Glu435Cys/Ser658Cys of BamA were cultured overnight in LB medium with 50 µg ml<sup>-1</sup> kanamycin, 100 µg ml<sup>-1</sup> carbenicillin and 0.025% L-(+)-arabinose, respectively. The membrane fraction from 50 ml cells was isolated and solubilized as described above. The samples were mixed with SDS loading buffer and then boiled for 5 min or kept at room temperature for 5–10 min. SDS-PAGE was performed at 4 °C by running the gel for 60 min at 150 V. The proteins were transferred to PVDF membrane as described above and the BamA mutants were detected by western blotting.

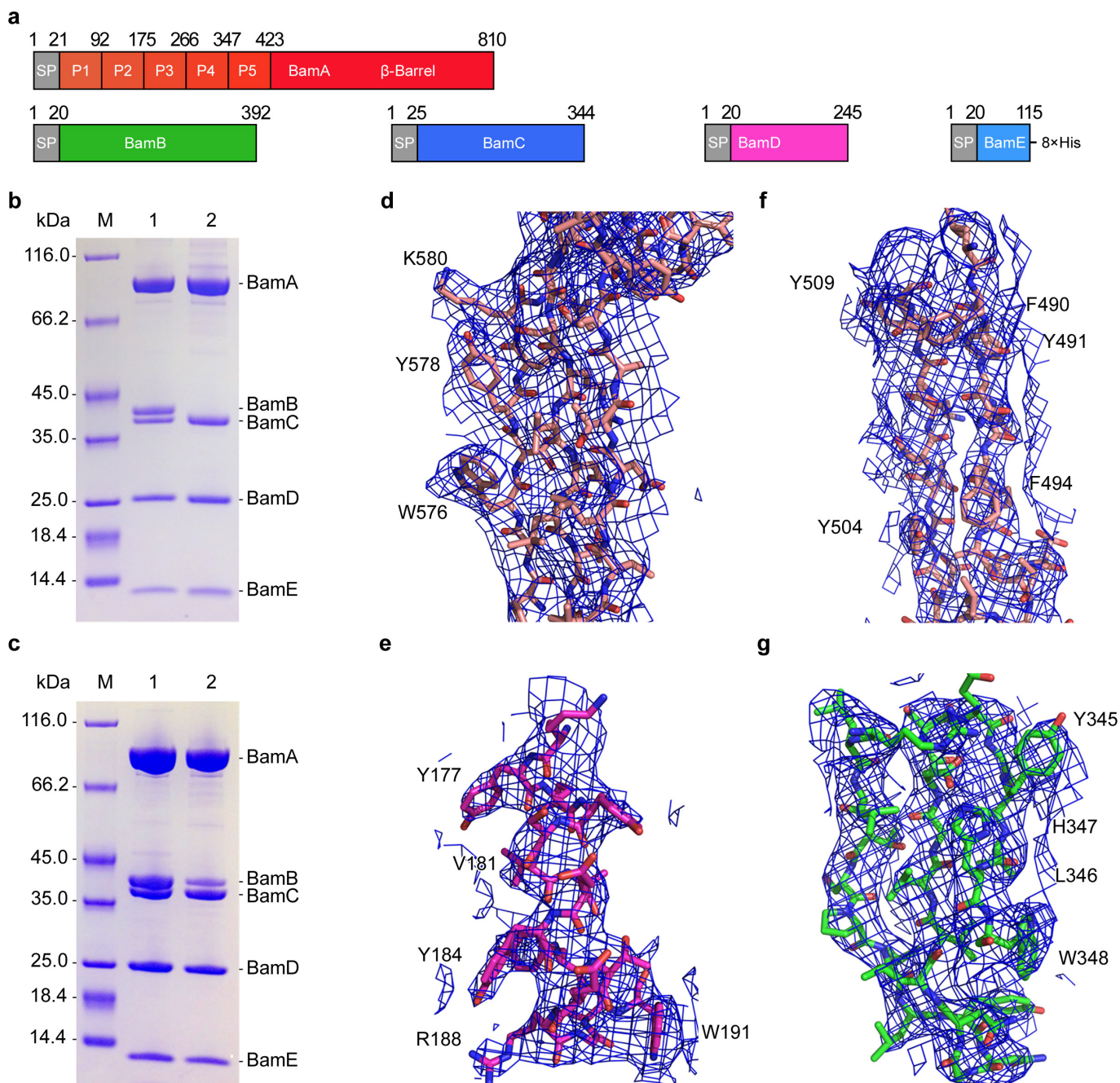
**Molecular modelling and simulations.** All molecular dynamics simulations were performed using GROMACS v5.0.2 (ref. 50). The Martini 2.2 force field<sup>51</sup> was used to run an initial 1 µs Coarse Grained (CG) molecular dynamics simulation to permit the assembly and equilibration of a 1-palmitoyl, 2-*cis*-vacenyl, phosphatidylglycerol (PVPG): 1-palmitoyl, 2-*cis*-vacenyl, phosphatidylethanolamine (PVPE) bilayers around the BamABCDE complexes<sup>52</sup>. Using the self-assembled system as a guide the coordinates of the BAM complexes were inserted into an asymmetric model *E. coli* OM, comprised of PVPE, PVPG, cardiolipin in the periplasmic leaflet and the inner core of Rd1 LPS lipids in the outer leaflet<sup>53</sup>, using Alchembed<sup>54</sup>. This equated to a total system size of ~500,000 atoms. The systems were then equilibrated for 1 ns with the protein restrained before 100 ns of unrestrained atomistic molecular dynamics using the Gromos53a6 force field<sup>55</sup>. The lipid-modified cysteine parameters were created from lipid parameters for diacylglycerol and palmitoyl and appended to the parameters of the N-terminal cysteines<sup>56</sup>. Systems were neutralised with Mg<sup>2+</sup> ions, to preserve the integrity of the outer leaflet of the OM, and a 150 mM concentration of NaCl.

All ~500,000 atom systems were all run for 100 ns, with box dimensions in the region of 200 × 200 × 150 Å<sup>3</sup>. To assess the stability of the subunit stoichiometry we assessed various combinations of BAM assemblies. For both BamACDE and BamABCDE crystal structures, we investigated ABCDE, AD and A alone, with three repeats each; while single simulations were also performed for BamABD, ACD, ADE, ABDE and ACDE, with a total simulation time equating to 2.8 µs. In cases where domains or subunits were missing these were added to the complex by structurally aligning the resolved units from the companion structure. For BamB, this was added to the BamACDE complex by structurally aligning POTRA 3. For the full BamC, this was added to the BamABCDE by aligning the resolved N-terminal domains. Individual protein complexes were configured and built using Modeller<sup>57</sup> and PyMOL (The PyMOL Molecular Graphics System, version 1.8, Schrödinger, LLC). All simulations were performed at 37 °C, with protein, lipids and solvent separately coupled to an external bath, using the velocity-rescale thermostat<sup>58</sup>. Pressure was maintained at 1 bar, with a semi-isotropic compressibility of 4 × 10<sup>-5</sup> using the Parrinello–Rahman barostat<sup>59</sup>. All bonds were constrained with the LINCS algorithm<sup>60,61</sup>. Electrostatics was measured using the Particle Mesh Ewald (PME) method<sup>62</sup>, while a cut-off was used for Lennard–Jones parameters, with a Verlet cut-off scheme to permit GPU calculation of non-bonded contacts. Simulations were performed with an integration time-step of 2 fs.

The linear interpolation between the three structures was performed using the morph operation in Gromacs tools<sup>50</sup>. Analysis of the molecular simulations was performed using Gromacs tools<sup>50</sup>, MDAnalysis<sup>63</sup> and locally written scripts.

Conservation analysis was performed using ConSurf<sup>64</sup>. For each subunit, 150 homologues were collected from UNIREF90<sup>65</sup> using three iterations of CSI-Blast<sup>66</sup>, with an E-value of 0.0001. The ConSurf scores were then mapped into the B-factor column for each of the subunits.

42. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature Methods* **4**, 251–256 (2007).
43. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
44. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010).
45. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
46. Cowtan, K. DM: An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34–38 (1994).
47. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
48. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
49. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).
50. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
51. de Jong, D. H. *et al.* Improved parameters for the Martini coarse-grained protein force field. *J. Chem. Theory Comput.* **9**, 687–697 (2013).
52. Stansfeld, P. J. *et al.* MemProtMD: Automated insertion of membrane protein structures into explicit lipid membranes. *Structure* **23**, 1350–1361 (2015).
53. Piggot, T. J., Pineiro, A. & Khalid, S. Molecular dynamics simulations of phosphatidylcholine membranes: a comparative force field study. *J. Chem. Theory Comput.* **8**, 4593–4609 (2012).
54. Jefferys, E., Sands, Z. A., Shi, J. Y., Sansom, M. S. P. & Fowler, P. W. Alchembed: a computational method for incorporating multiple proteins into complex lipid geometries. *J. Chem. Theory Comput.* **11**, 2743–2754 (2015).
55. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25**, 1656–1676 (2004).
56. Domański, J., Stansfeld, P. J., Sansom, M. S. P. & Beckstein, O. Lipidbook: a public repository for force-field parameters used in membrane simulations. *J. Membr. Biol.* **236**, 255–258 (2010).
57. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **15**, 5.6.1–5.6.30 (2006).
58. Bussi, G., Zykova-Timan, T. & Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **130**, 074101 (2009).
59. Parrinello, M. & Rahman, A. Polymorphic transitions in single-crystals - a new molecular-dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
60. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
61. Hess, B. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–437 (2008).
62. York, D. M., Wlodawer, A., Pedersen, L. G. & Darden, T. A. Atomic-level accuracy in simulations of large protein crystals. *Proc. Natl Acad. Sci. USA* **91**, 8715–8718 (1994).
63. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. Software news and updates MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
64. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533 (2010).
65. Supek, B. *et al.* UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).
66. Biegert, A. & Soding, J. Sequence context-specific profiles for homology searching. *Proc. Natl Acad. Sci. USA* **106**, 3770–3775 (2009).

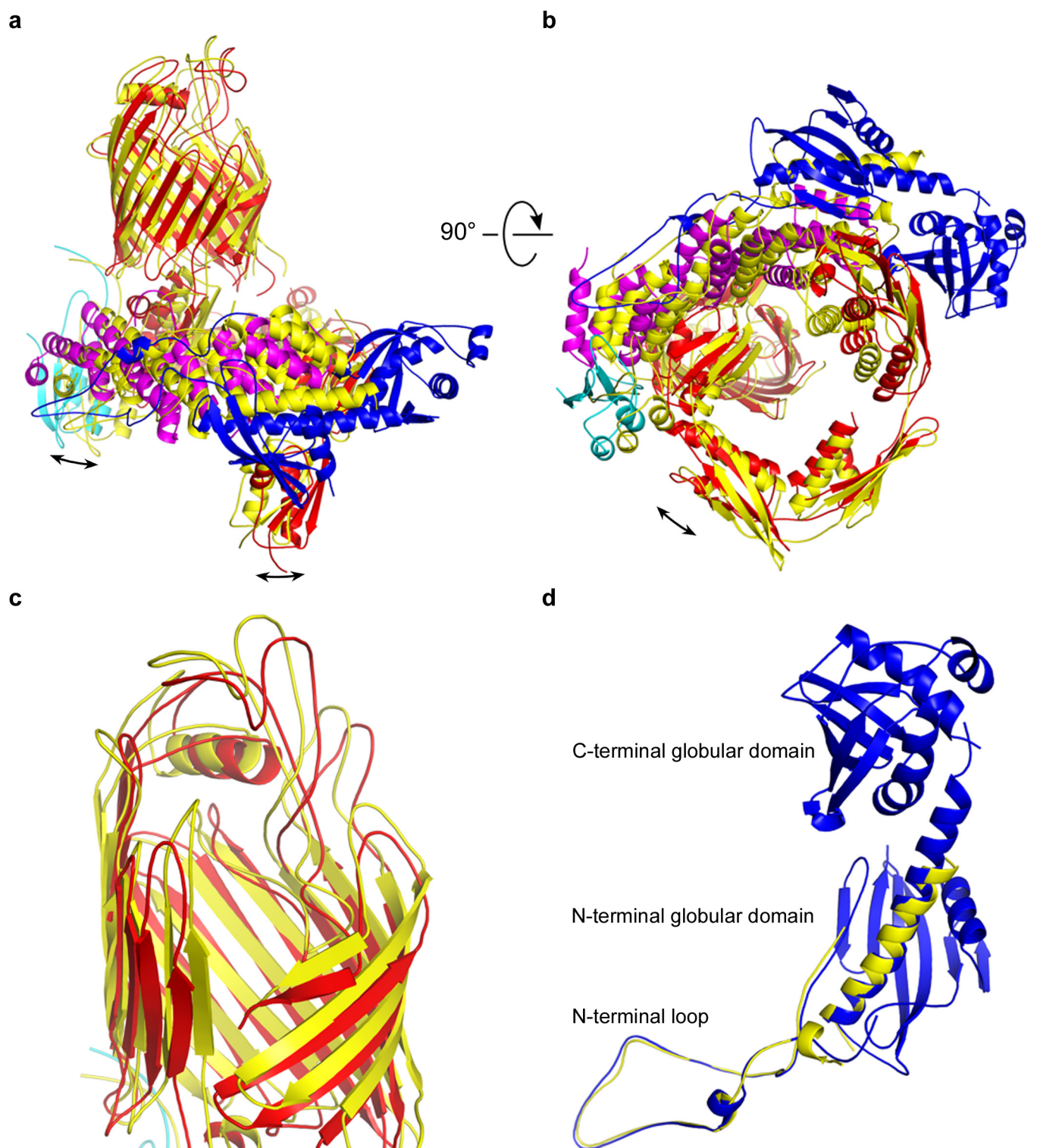


**Extended Data Figure 1 | BamABCDE and BamACDE complexes and electron density maps contoured at  $1\sigma$ .** **a**, Schematic diagram of the five BAM subunits. P-1 to P-5 represent the five BamA POTRA domains.

**b**, SDS-PAGE analysis of the BAM complex from crystals. M, molecular mass marker; 1 and 2 are crystals of the purified BAM complex expressed by construct pYG120 and pJH114, respectively (Supplementary Fig. 1). The BamABCDE crystals contain the full-length BamA–BamE. The crystals were washed five times in fresh reservoir solution, and then dissolved in SDS-PAGE loading buffer. The results showed that the BamB is absent in the BamACDE crystals, while the BamC is complete in both

the BamABCDE and BamACDE crystals. **c**, SDS-PAGE analysis of the purified BAM complex. The BAM complexes expressed from pJH114 is a mixture of BamABCDE and BamACDE complexes (Supplementary Fig. 1). **d**,  $2F_o - F_c$  electron density map of BamA residues Trp576–Lys580 of BamACDE contoured at  $1\sigma$ . **e**,  $2F_o - F_c$  electron density map of BamD residues Tyr177–Trp191 of BamACDE contoured at  $1\sigma$ . **f**,  $2F_o - F_c$  electron density map of BamA residues Tyr504–Tyr509 and Phe490–Phe494 of BamABCDE complex contoured at  $1\sigma$ . **g**,  $2F_o - F_c$  electron density map of BamB residues Tyr345–Trp348 contoured at  $1\sigma$ .

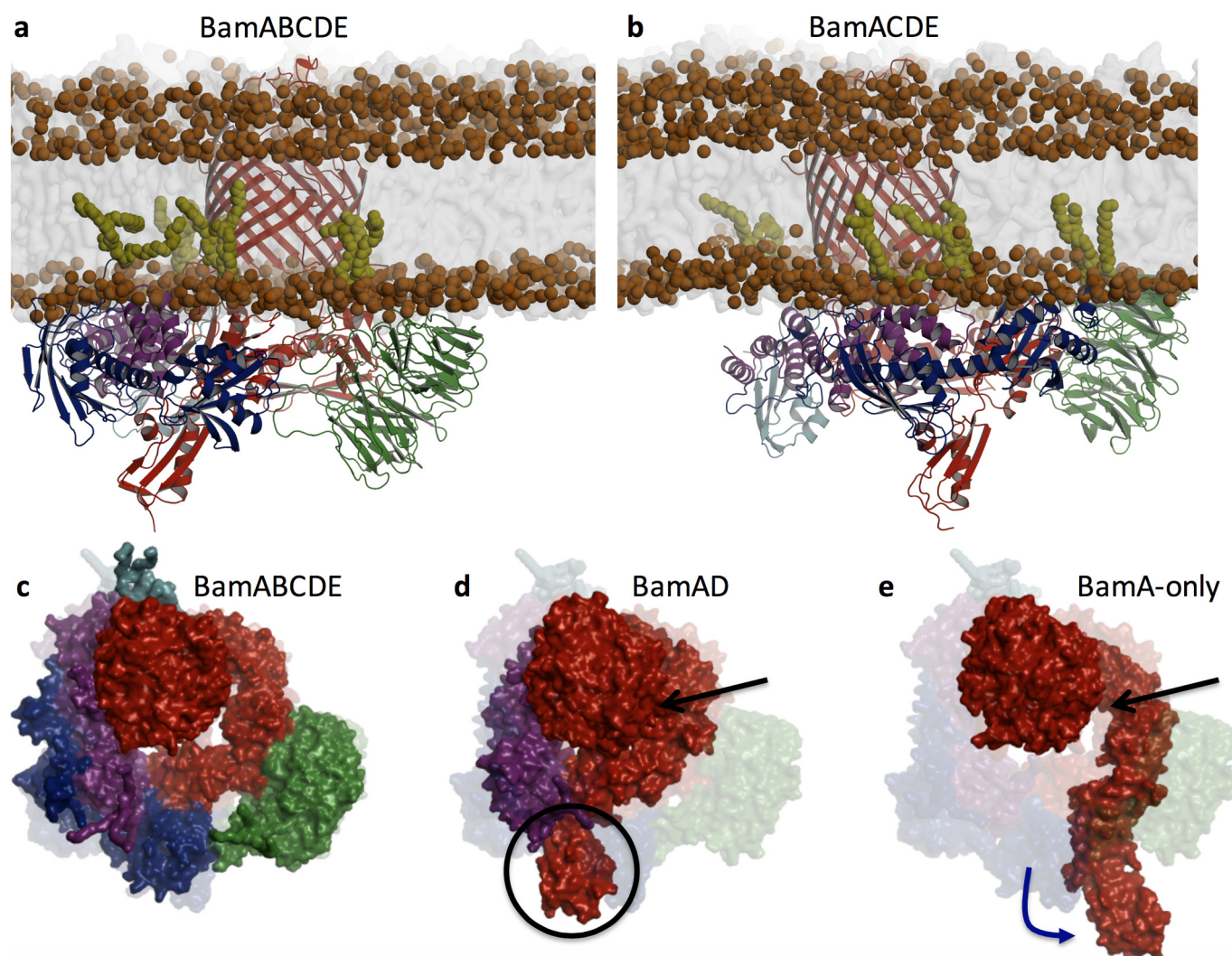




**Extended Data Figure 2 | Superimposition of the two BamACDE complexes in the asymmetric unit.** The BamACDE complex with the full-length BamC, showing BamA (red), BamC (blue), BamD (magenta) and BamE (cyan). Only N-terminal loop of BamC was observed in another BamACDE complex in the asymmetric unit cell (yellow). The structure data suggests that the role of BamC is to retain the ring structure of BamA and BamD during OMP insertion. **a**, Membrane view of the superimposed BamACDE complexes. The primary difference is one complex has a complete BamC subunit, which binds BamD, BamE, POTRA 1 and 2, while the second complex only the N-terminal coil structure up to Pro88

is observed and the rest of BamC is disordered. The overall structures of the two complexes are very similar with some conformational changes in the  $\beta$ -strands of barrel and extracellular loops with r.m.s.d. of 0.908 Å over 378 C $\alpha$  atoms, while the periplasmic circular structure has some rotation (see arrows) with a r.m.s.d. of 4.706 Å over 385 C $\alpha$  atoms. **b**, Periplasmic view of the superimposition of the two structures. The periplasmic circular structure has some rotations when the C-terminal global domain binds on the POTRA 2. **c**, Superimposition of the barrels of the two complexes. **d**, Superimposition of the two BamCs. The N-terminal coil structures superimpose well with a r.m.s.d. of 0.807 Å over 86 C $\alpha$  atoms.





**Extended Data Figure 3 | Molecular dynamics simulation of BAM complexes.** **a, b**, BamABCDE (**a**) and BamACDE (**b**) structures modelled with all subunits present and embedded in a model *E. coli* outer membrane (grey). Phosphate atoms are shown in orange spheres. Lipid-modified cysteine residues of BamB, BamC, BamD and BamE are shown in yellow spheres. **c**, Both complexes are stable in molecular dynamics simulations, showing limited deviation from the starting configuration (shown in the background). **d**, Simulations of the complexes of only BamA and BamD subunits retain the ring structure. Without BamC present POTRA 1

(black circle) moves towards the membrane, while POTRA 3 (black arrow) moves towards and interacts with the periplasmic loops of the barrel. The dynamics of POTRA 3 appear to be modulated by BamB. **e**, Simulations of BamA show enhanced dynamics of the POTRA domains, with POTRA 1 and 2 rotating towards the membrane in an anti-clockwise direction (blue arrow). This separates POTRA 3 from the barrel (black arrow). This conformation of the POTRA domains is unable to form the BAM ring, highlighting the essential nature of BamD and its interactions with BamA in maintaining the ring structure.

a



b



c



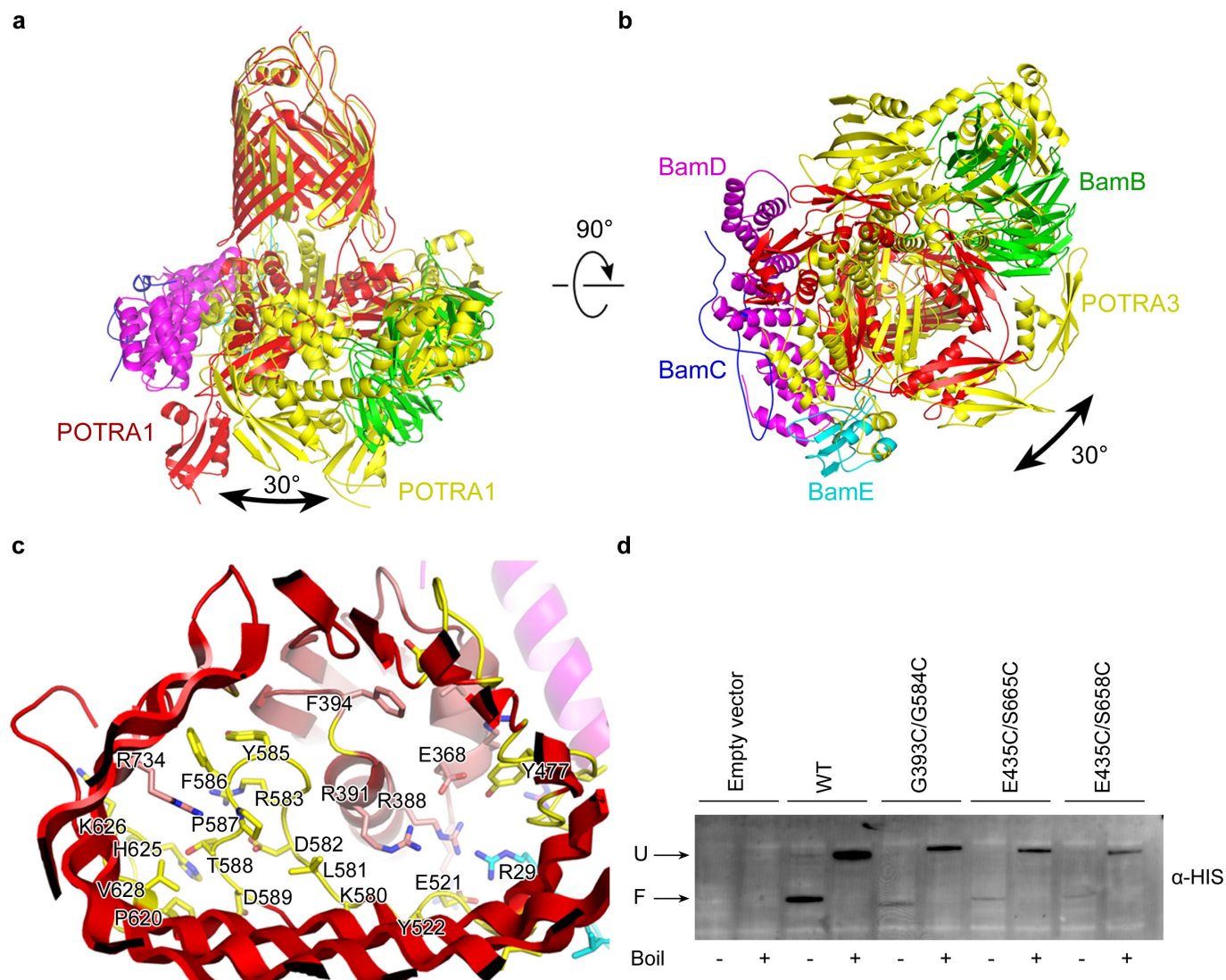
d



**Extended Data Figure 4 | BamA of the BamABCDE complex is superimposed onto the other published BamA structures.** All the published BamA structures are in the inward-open conformation. In all cases the BamA from BamABCDE is shown in red. **a**, The BamA of BamABCDE complex is superimposed onto BamA of *N. gonorrhoeae* (grey) (PDB accession 4K3B)<sup>18</sup>. The two barrel structures are similar with a r.m.s.d. of 3.803 Å over 385 C $\alpha$  atoms, but the conformations of the five POTRA domains are quite different. The dotted circle indicates the hydrophobic gap between  $\beta$ 1C and  $\beta$ 15C. **b**, BamA of *E. coli* (magenta)

(PDB accession 4N75)<sup>19</sup>. The two barrel structures superimpose well with a r.m.s.d. of 0.644 Å over 385 C $\alpha$  atoms, but differences are observed for the  $\beta$ 16C terminal residues. The C-terminal residues in BamA of BamABCDE move towards the lumen of the barrel. **c**, BamA of *E. coli* (yellow) (PDB accession 4C4V)<sup>20</sup> with a r.m.s.d. of 1.382 Å over 365 barrel C $\alpha$  atoms. The conformations of the POTRA 5 are quite different. **d**, BamA of *Haemophilus ducreyi* (green) (PDB accession 4K3C)<sup>18</sup>. The barrel structures are similar with a r.m.s.d. of 2.376 Å over 365 barrel C $\alpha$  atoms, but the conformations of POTRA 4 and 5 are quite different.

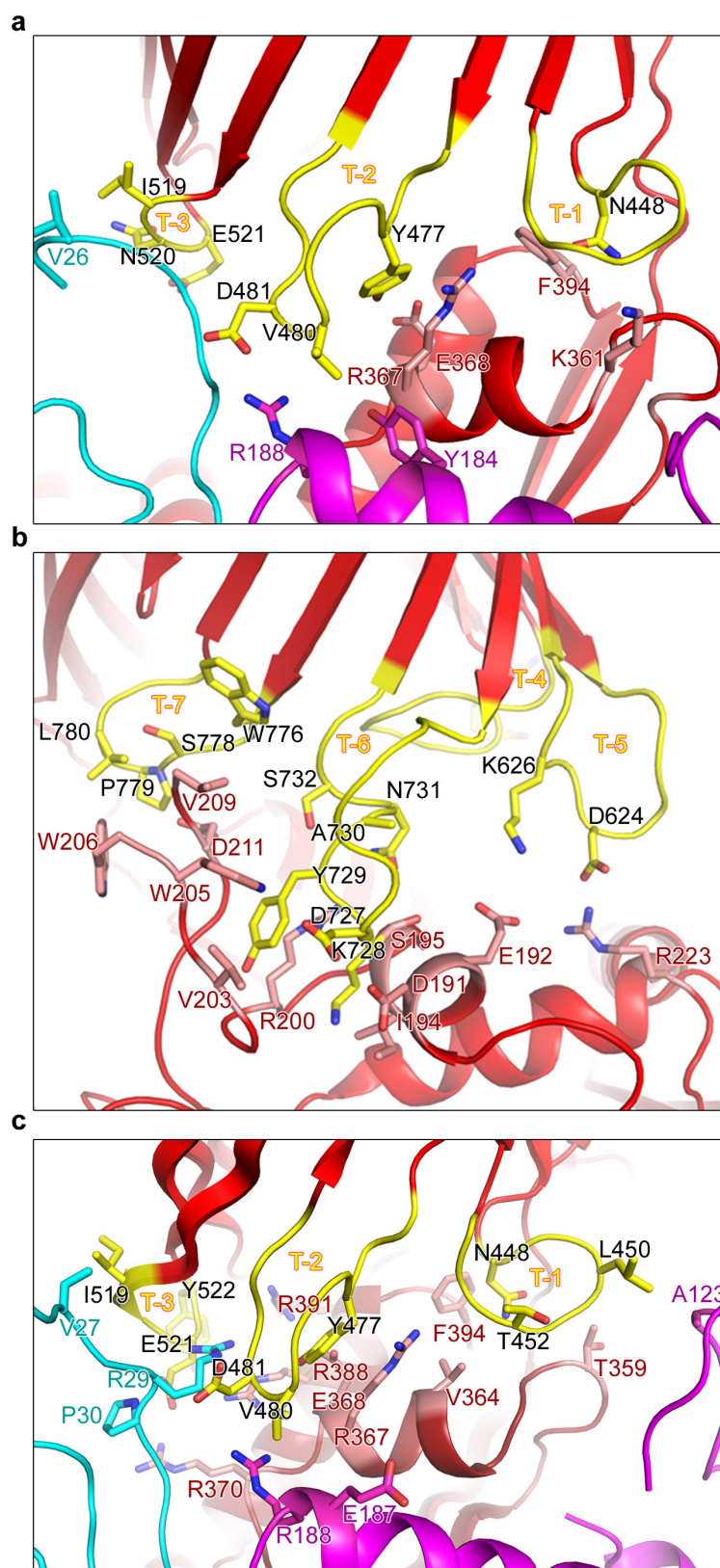




**Extended Data Figure 5 | The conformational changes between the BamABCDE and BamACDE complexes and heat-modifiability assays of the BamA double cysteine mutants.** The two structures are superimposed onto the BamA barrel structures of BamABCDE and BamACDE complexes with a r.m.s.d. of 4.85 Å over the 379 barrel C $\alpha$  atoms and a maximum r.m.s.d. of 20 Å. The POTRA domains align with an r.m.s.d. of 5.764 Å over 384 C $\alpha$  atoms with maximum 15 Å. The BamABCDE complex is in the same colour scheme as Fig. 1. The BamACDE complex is in yellow. The barrel strands  $\beta$ 1C– $\beta$ 6C rotate around 65° from BamABCDE to BamACDE, while the BAM periplasmic unit rotates around 30° in a anti-clockwise direction from BamABCDE to BamACDE. **a**, Membrane view of the superimposition of the BamABCDE

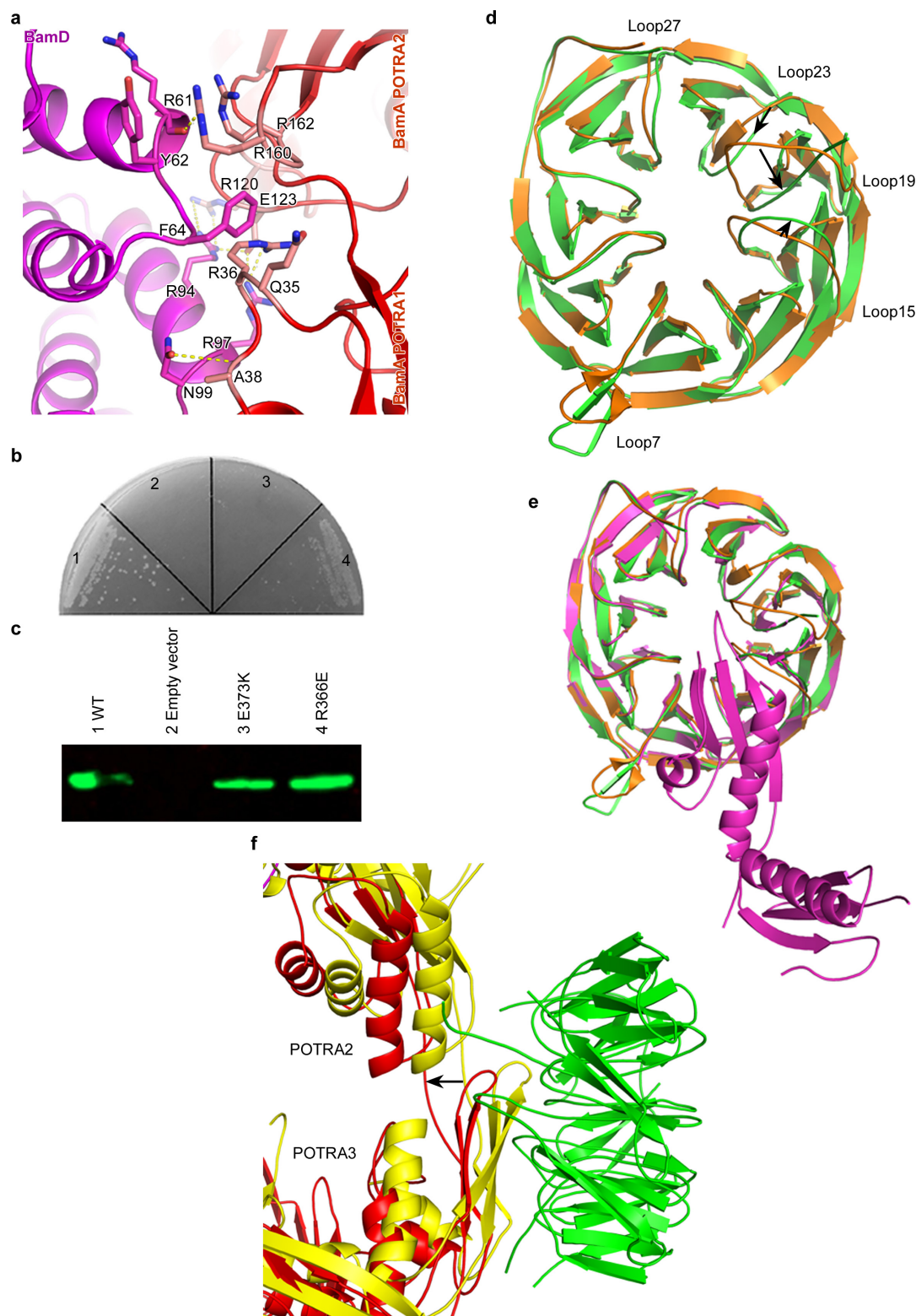
and BamACDE complexes. The conformations of BamA POTRA domains, BamB, BamC and BamD are considerably different between the two complexes. **b**, The periplasmic view of the superimposition of BamABCDE and BamACDE. The circular units rotate around 30° between the two BAM complexes. **c**, The residues involved in closing the barrel at the periplasmic side in the BamACDE structure. **d**, Heat-modifiability assays of the BamA double cysteine mutants. SDS-PAGE/western blot analysis of the wild-type BamA, BamA Gly393Cys/Gly584Cys, Glu435Cys/Ser665Cys and Glu435Cys/Ser658Cys mutants showed the heat-modifiability, indicating that the three double cysteine BamA mutants were correctly folded into the OM. F, folded; U, unfolded. See Supplementary Fig. 5.





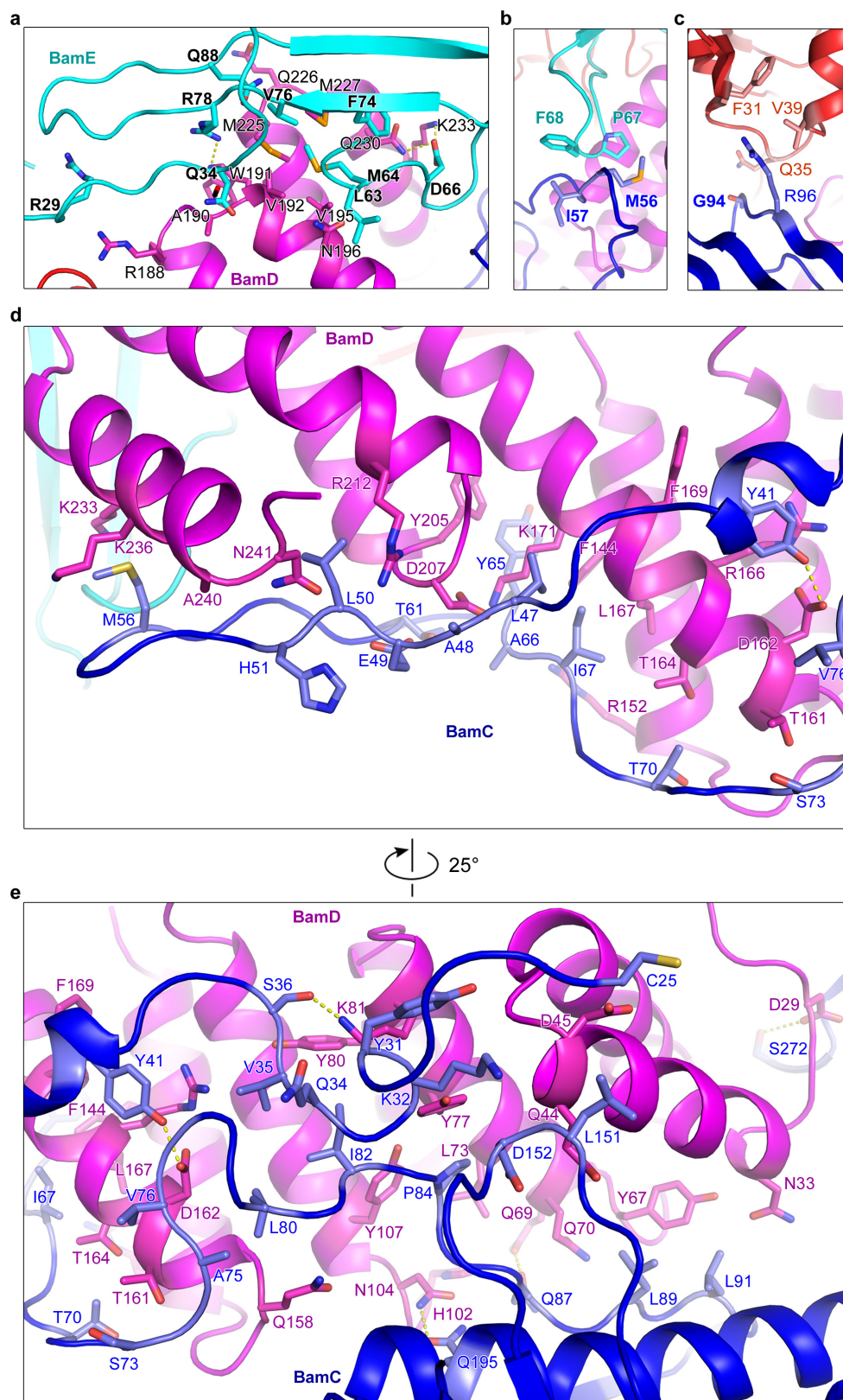
**Extended Data Figure 6 | Periplasmic loops bind to BamA POTRA 3, 5, BamD and BamE.** In the BamABCDE complex, the BamA barrel interacts with POTRA 3, 5, BamE and BamD through the periplasmic turns T-1, -2, -3, -5, -6 and -7. **a**, In the BamABCDE complex, the residues of T-1, -2 and -3 are involved in the interactions with POTRA 5, BamD and BamE. **b**, Residues in T-5, -6 and -7 interact with POTRA 3 in the BamABCDE

complex. **c**, In the BamACDE complex no interactions are observed between the periplasmic turns and POTRA 3. The figure shows that the residues in T-1, -2 and -3 interact with residues in POTRA 5, BamD and BamE. These structural data may suggest that BamB, C, D and E either directly or indirectly control the conformation of the barrel through its periplasmic turns.



**Extended Data Figure 7 | BamA and BamD interactions, and superimposition of the BamB structures and the conformational changes of POTRA2 and 3.** **a**, BamA POTRA 1 and 2 interact with the N-terminal domain of BamD. The interacting residues from both BamA and BamD are shown. **b**, The functional assays and the western blots were repeated at least three times. Functional assays of the BamA interaction with BamD. The mutation BamA Glu373Lys is lethal, while mutant Arg366Glu impairs the bacterial growth, suggesting these residues may have an important role in the BAM complex. 1–4 are as shown in **c**. **c**, Protein expression levels of BamA mutations were detected by western blotting. **d**, Periplasmic view of BamB of the BamABCDE complex (green)

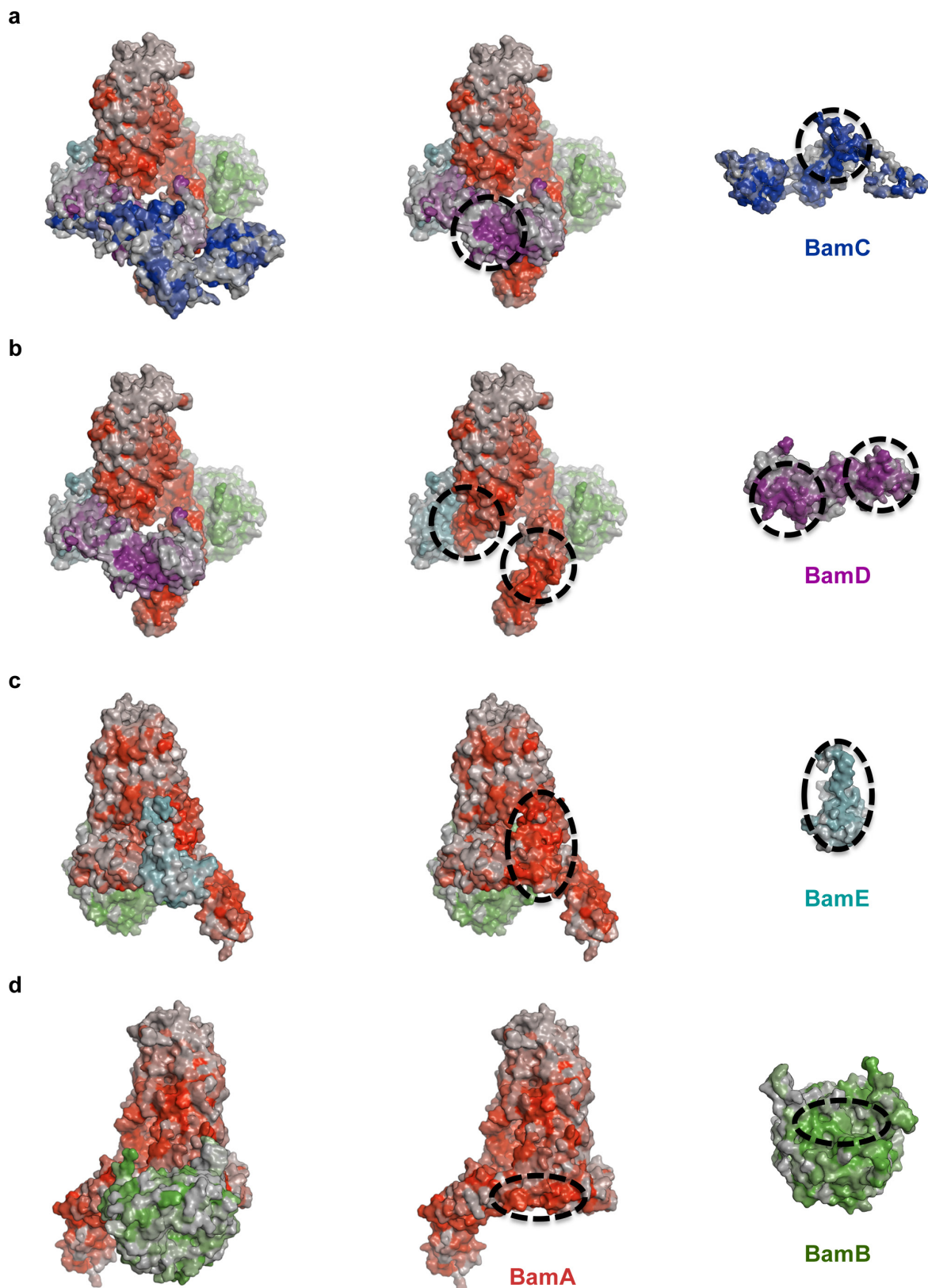
superimposed onto the free BamB structure (orange) (PDB code 3Q7N)<sup>21</sup> with a r.m.s.d. of 1.81 Å over 351 C $\alpha$  atoms with the maximum deviation of 12 Å at loop 19. Loops 15, 19, 23 and 27 of BamB adopt conformational changes to bind to POTRA 2 and 3. **e**, BamB of the BamABCDE complex superimposed onto BamB in complex with POTRA 3 and 4 (magenta) (PDB code 4PK1)<sup>32</sup>. The two BamB structures are very similar with a r.m.s.d. of 0.5860 Å over 341 C $\alpha$  atoms. **f**, Superimposition of BamABCDE and BamACDE at POTRA 2 and 3 with a r.m.s.d. of 3.57 Å over 159 C $\alpha$  atoms. In the BamACDE structure the hinge angle between POTRA 2 and 3 is reduced, while POTRA 2 and 3 also separate from BamB, reducing the interactions between BamB, and POTRA 2 and 3.



**Extended Data Figure 8 | BamE interacts with BamD and BamC, and BamC interactions with BamD.** BamE interacts with BamA, BamD and BamC. BamC binds extensively to the C-terminal domain of BamD. **a**, BamE interacts with BamD in the BamACDE complex. BamE contacts the C-terminal domain residues of BamD in the BamACDE complex. **b**, BamE forms hydrophobic interactions with BamC in the BamACDE

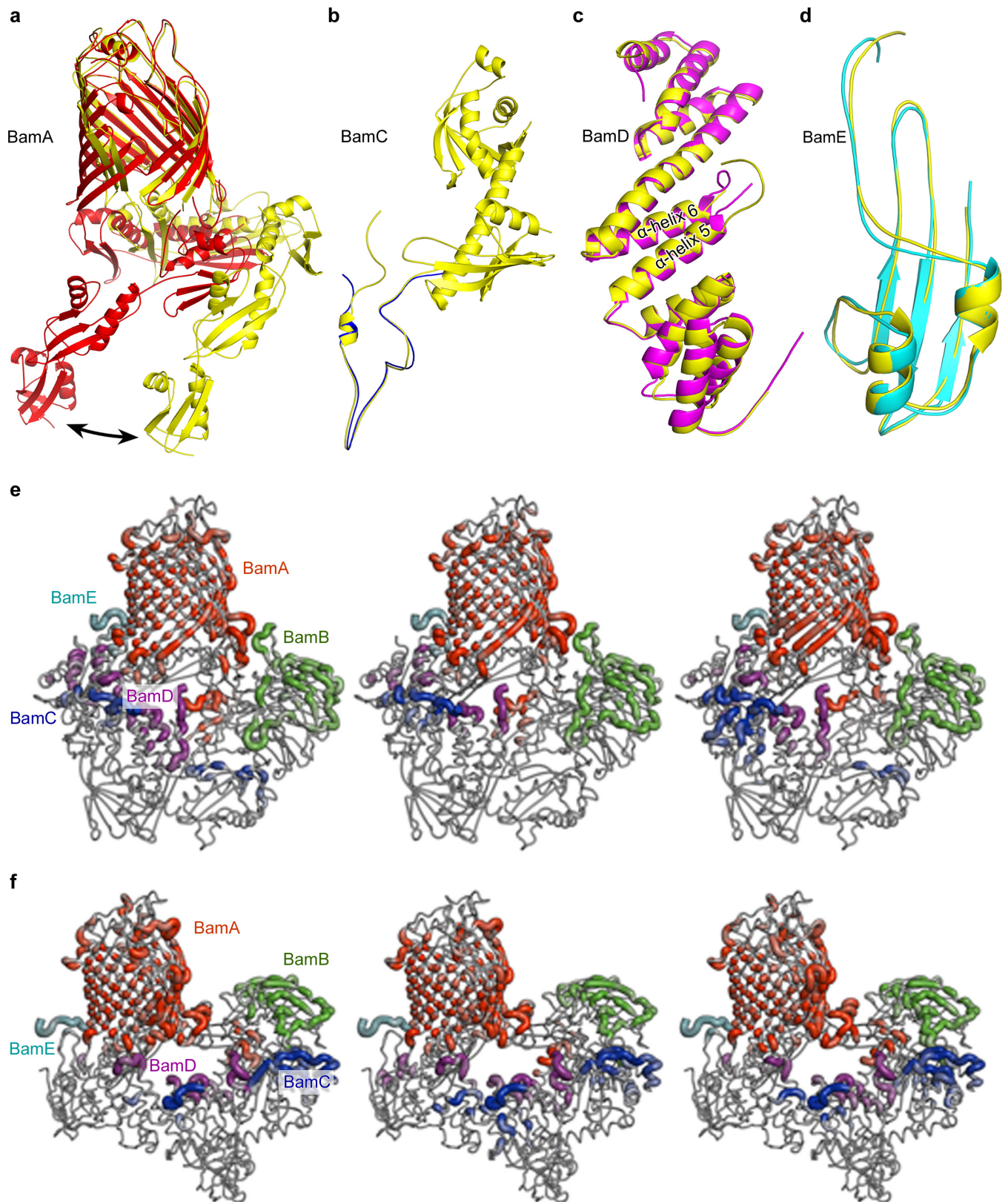
complex. BamE residues Pro67, Phe68 and BamC residues Met56 and Ile57 are shown. **c**, BamC forms contacts with BamA POTRA 1 in BamACDE. BamA residues Phe31, Gln35, Val39 and BamC residues Gly94 and Arg96 are shown. **d**, BamC interacts with the C-terminal domain of BamD. The interacting residues are shown as sticks. **e**, BamC interacts with the N-terminal domain of BamD.





**Extended Data Figure 9 | Conserved residues analysis of BAM complex.**  
**a–d**, Consurf residue conservation scores (1–9), plotted onto the molecular surfaces as a colour scale for BamA (red), BamB (green), BamC (blue), BamD (purple) and BamE (cyan), for the BamABCDE structure. Regions of white/grey indicate poorly conserved residues, whereas a more

intense colour indicates highly conserved residues. Black dashed circles represent the interaction points on removal of BamC (**a**), BamD (**b**), BamE (**c**) and BamB (**d**). For each interaction patch a high density of conserved residues is apparent.



Extended Data Figure 10 | See next page for caption.

**Extended Data Figure 10 | Conformational differences of the BAM subunits between the BamABCDE and BamACDE complexes, and BAM complex interacts with lipid of the OM.** The subunits of BamABCDE are coloured as in Fig. 1, while the BamACDE subunits are in yellow.

**a**, Superimposition of the BamA subunits onto the barrel domain with an r.m.s.d. of 4.85 Å over the 379 barrel C $\alpha$  atoms and an r.m.s.d. of 5.76 Å over the 384 C $\alpha$  atoms of the POTRA domains. The BamA barrel has notable conformational changes in  $\beta$ 1C– $\beta$ 6C. The periplasmic POTRA domains rotate about 30° from BamABCDE complex to BamACDE complex, suggesting a novel rotation mechanism to facilitate OMP insertion into the OM. **b**, Superimposition of the BamC structures. The BamC structures have some conformational changes with a r.m.s.d. of 2.102 Å over 47 C $\alpha$  atoms of the BamC N-terminal loop. The N-terminal loop Cys25–Val35 becomes more ordered in BamACDE complex. Particularly, the N-terminal domain and the C-terminal domain are ordered and bind to POTRA 1, 2 and the N-terminal domain of BamD in BamACDE complex. The N-terminal loops of the BamC structures superimpose well between residues Val35 and Pro88. **c**, Superimposition of

the BamD structures with an r.m.s.d. of 1.201 Å over 203 C $\alpha$  atoms. The  $\alpha$ -helices are conserved, but the loops have some conformational changes, especially loop 6 (residues Asp121–Asp136) between  $\alpha$ -helix 5 and  $\alpha$ -helix 6. **d**, Superimposition of BamE structures with a r.m.s.d. of 1.721 Å over 81 C $\alpha$  atoms. The  $\beta$ -strands and  $\alpha$ -helices of BamE are well conserved, with minor conformational changes observed in the loops. **e**, Lipid–protein interactions for the BamACDE structure. BamB was modelled into the BamACDE complex by molecular modelling. The BamABCDE complex was built in the OM (Methods), and the residues interacting with lipids of the OM with 4 Å are shown in putty representation to depict lipid interaction residues. Equivalent residues in all five subunits BamA (red), BamB (green), BamC (blue), BamD (purple) and BamE (cyan) interact with the membrane in all three independent simulations. **f**, Lipid–protein interactions for the BamABCDE structure. BamC was added to the BamABCDE complex by molecular modelling, using the solved domain from the companion complex. BamABCDE complex was inserted into the OMP, with lipid anchors designed (Methods).



Extended Data Table 1 | Data collection and refinement statistics

	BamACDE Se-Met <sup>‡a</sup>	BamABCDE NaI <sup>a</sup>	BamABCDE Native <sup>‡a</sup>
<b>Data collection</b>			
Space group	<i>P</i> 4 <sub>2</sub> 2 <sub>1</sub> 2	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2	<i>P</i> 4 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	254.16, 254.16, 179.22	116.72, 116.72, 432.44	116.69, 116.69, 435.19
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.97951	1.82330	0.97949
Resolution (Å)	29.94–3.90 (4.02–3.90)*	29.86–4.00 (4.27–4.00)	49.65–2.90 (2.97–2.90)
<i>R</i> <sub>merge</sub> (%)	28.5 (>100.0)	24.8 (>100.0)	18.0 (>100)
<i>CC</i> 1/2 (%)	99.9 (49.4)	100 (99.6)	99.8 (50.8)
<i>I</i> / $\sigma$ <i>I</i>	11.0 (0.9)	37.0 (11.8)	15.0 (0.6)
Completeness (%)	99.8 (100.0)	98.5 (97.8)	100 (100)
Redundancy	27.1 (27.2)	158.00 (165.1)	26.4(23.8)
<b>Refinement</b>			
Resolution (Å)	29.92 – 3.90		49.65 – 2.90
No. reflections	73745		67553
<i>R</i> <sub>factor</sub> / <i>R</i> <sub>free</sub>	30.44/31.93		27.62/30.41
No. atoms			
Protein	19796		22815
Ligand/ion	0		0
Water	0		0
<i>B</i> -factors(Å <sup>2</sup> )			
Protein	150		118
Ligand/ion	N/A		N/A
Water	N/A		N/A
R.m.s. deviations			
Bond lengths (Å)	0.010		0.009
Bond angles (°)	1.868		1.609
Residues in			
Ramachandran plot			
Favored (%)	90.5		91.6
Allowed (%)	8.7		7.7
Outliers (%)	0.8		0.7
PDB code	5D0Q		5D0O

\*Values in parentheses are for highest-resolution shell.

<sup>‡</sup> Highest resolution shell was taken as point where *CC*1/2 > 30 along strongest reciprocal lattice direction.

<sup>a</sup>Data statistics shown for each wavelength are a combination of two datasets (BamACDE Se-Met) and four datasets (BamABCDE NaI).

<sup>b</sup>*R*<sub>factor</sub> =  $\Sigma||F_o| - |F_c|| / \Sigma|F_o|$ , where *F*<sub>o</sub> and *F*<sub>c</sub> are observed and calculated as structure factors, respectively.

<sup>c</sup>*R*<sub>free</sub> is calculated using 5% of total reflections, which is randomly selected as a free group and not used in refinement.

Diffraction data for all structures were anisotropic and axis specific resolution cutoffs from AIMLESS (*CC*1/2>0.3) for refinement data basis are listed below for illustration:

BamACDE Se-Met      h-k plane = 4.45 Å, l axis = 3.50 Å

BamABCDE Native      h-k plane = 3.48 Å, l axis = 2.75 Å

# A large light-mass component of cosmic rays at $10^{17}$ – $10^{17.5}$ electronvolts from radio observations

S. Buitink<sup>1,2</sup>, A. Corstanje<sup>2</sup>, H. Falcke<sup>2,3,4,5</sup>, J. R. Hörandel<sup>2,4</sup>, T. Huege<sup>6</sup>, A. Nelles<sup>2,7</sup>, J. P. Rachen<sup>2</sup>, L. Rossetto<sup>2</sup>, P. Schellart<sup>2</sup>, O. Scholten<sup>8,9</sup>, S. ter Veen<sup>3</sup>, S. Thoudam<sup>2</sup>, T. N. G. Trinh<sup>8</sup>, J. Anderson<sup>10</sup>, A. Asgekar<sup>3,11</sup>, I. M. Avruch<sup>12,13</sup>, M. E. Bell<sup>14</sup>, M. J. Bentum<sup>3,15</sup>, G. Bernardi<sup>16,17</sup>, P. Best<sup>18</sup>, A. Bonafede<sup>19</sup>, F. Breitling<sup>20</sup>, J. W. Broderick<sup>21</sup>, W. N. Brouw<sup>3,13</sup>, M. Brüggen<sup>19</sup>, H. R. Butcher<sup>22</sup>, D. Carbone<sup>23</sup>, B. Ciardi<sup>24</sup>, J. E. Conway<sup>25</sup>, F. de Gasperin<sup>19</sup>, E. de Geus<sup>3,26</sup>, A. Deller<sup>3</sup>, R.-J. Dettmar<sup>27</sup>, G. van Diepen<sup>3</sup>, S. Duscha<sup>3</sup>, J. Eislöffel<sup>28</sup>, D. Engels<sup>29</sup>, J. E. Enriquez<sup>3</sup>, R. A. Fallows<sup>3</sup>, R. Fender<sup>30</sup>, C. Ferrari<sup>31</sup>, W. Frieswijk<sup>3</sup>, M. A. Garrett<sup>3,32</sup>, J. M. Grießmeier<sup>33,34</sup>, A. W. Gunst<sup>3</sup>, M. P. van Haarlem<sup>3</sup>, T. E. Hassall<sup>21</sup>, G. Heald<sup>3,13</sup>, J. W. T. Hessels<sup>3,23</sup>, M. Hoefl<sup>28</sup>, A. Horneffer<sup>5</sup>, M. Iacobelli<sup>3</sup>, H. Intema<sup>32,35</sup>, E. Juette<sup>27</sup>, A. Karastergiou<sup>30</sup>, V. I. Kondratiev<sup>3,36</sup>, M. Kramer<sup>5,37</sup>, M. Kuniyoshi<sup>38</sup>, G. Kuper<sup>3</sup>, J. van Leeuwen<sup>3,23</sup>, G. M. Looze<sup>3</sup>, P. Maat<sup>3</sup>, G. Mann<sup>20</sup>, S. Markoff<sup>23</sup>, R. McFadden<sup>3</sup>, D. McKay-Bukowski<sup>39,40</sup>, J. P. McKean<sup>3,13</sup>, M. Mevius<sup>3,13</sup>, D. D. Mulcahy<sup>21</sup>, H. Munk<sup>3</sup>, M. J. Norden<sup>3</sup>, E. Orru<sup>3</sup>, H. Paas<sup>41</sup>, M. Pandey-Pommier<sup>42</sup>, V. N. Pandey<sup>3</sup>, M. Pietka<sup>30</sup>, R. Pizzo<sup>3</sup>, A. G. Polatidis<sup>3</sup>, W. Reich<sup>5</sup>, H. J. A. Röttgering<sup>32</sup>, A. M. M. Scaife<sup>21</sup>, D. J. Schwarz<sup>43</sup>, M. Serylak<sup>30</sup>, J. Sluman<sup>3</sup>, O. Smirnov<sup>17,44</sup>, B. W. Stappers<sup>3</sup>, M. Steinmetz<sup>20</sup>, A. Stewart<sup>30</sup>, J. Swinbank<sup>23,45</sup>, M. Tagger<sup>33</sup>, Y. Tang<sup>3</sup>, C. Tasse<sup>44,46</sup>, M. C. Toribio<sup>3,32</sup>, R. Vermeulen<sup>3</sup>, C. Vocks<sup>20</sup>, C. Vogt<sup>3</sup>, R. J. van Weeren<sup>16</sup>, R. A. M. J. Wijers<sup>23</sup>, S. J. Wijnholds<sup>3</sup>, M. W. Wise<sup>3,23</sup>, O. Wucknitz<sup>5</sup>, S. Yatawatta<sup>3</sup>, P. Zarka<sup>47</sup> & J. A. Zensus<sup>5</sup>

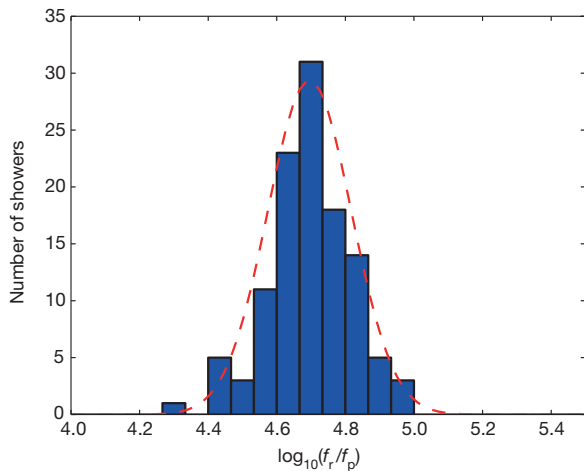
Cosmic rays are the highest-energy particles found in nature. Measurements of the mass composition of cosmic rays with energies of  $10^{17}$ – $10^{18}$  electronvolts are essential to understanding whether they have galactic or extragalactic sources. It has also been proposed that the astrophysical neutrino signal<sup>1</sup> comes from accelerators capable of producing cosmic rays of these energies<sup>2</sup>. Cosmic rays initiate air showers—cascades of secondary particles in the atmosphere—and their masses can be inferred from measurements of the atmospheric depth of the shower maximum<sup>3</sup> ( $X_{\max}$ ; the depth of the air shower when it contains the most particles) or of the composition of shower particles reaching the ground<sup>4</sup>. Current measurements<sup>5</sup> have either high uncertainty, or a low duty cycle and a high energy threshold. Radio detection of cosmic rays<sup>6–8</sup> is a rapidly developing technique<sup>9</sup> for determining  $X_{\max}$  (refs 10, 11) with a duty cycle of, in principle, nearly 100 per cent. The radiation is generated by the separation of relativistic electrons and positrons in the geomagnetic field and a negative charge excess in the shower front<sup>6,12</sup>. Here we report radio measurements of  $X_{\max}$  with a mean uncertainty of 16 grams per square centimetre for air showers

initiated by cosmic rays with energies of  $10^{17}$ – $10^{17.5}$  electronvolts. This high resolution in  $X_{\max}$  enables us to determine the mass spectrum of the cosmic rays: we find a mixed composition, with a light-mass fraction (protons and helium nuclei) of about 80 per cent. Unless, contrary to current expectations, the extragalactic component of cosmic rays contributes substantially to the total flux below  $10^{17.5}$  electronvolts, our measurements indicate the existence of an additional galactic component, to account for the light composition that we measured in the  $10^{17}$ – $10^{17.5}$  electronvolt range.

Observations were made with the Low Frequency Array (LOFAR<sup>13</sup>), a radio telescope consisting of thousands of crossed dipoles with built-in air-shower-detection capability<sup>14</sup>. LOFAR continuously records the radio signals from air showers, while simultaneously running astronomical observations. It comprises a scintillator array (LORA) that triggers the read-out of buffers, storing the full waveforms received by all antennas.

We selected air showers from the period June 2011 to January 2015 with radio pulses detected in at least 192 antennas. The total uptime was about 150 days, limited by construction and commissioning of the

<sup>1</sup>Astrophysical Institute, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. <sup>2</sup>Department of Astrophysics/IMAPP, Radboud University Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands. <sup>3</sup>ASTRON, Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA Dwingelo, The Netherlands. <sup>4</sup>Nikhef, Science Park Amsterdam, 1098 XG Amsterdam, The Netherlands. <sup>5</sup>Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany. <sup>6</sup>Institute for Nuclear Physics (IKP), Karlsruhe Institute of Technology (KIT), Postfach 3640, 76021 Karlsruhe, Germany. <sup>7</sup>Department of Physics and Astronomy, University of California Irvine, Irvine, California 92697, USA. <sup>8</sup>KVI Center for Advanced Radiation Technology, University of Groningen, 9747 AA Groningen, The Netherlands. <sup>9</sup>Vrije Universiteit Brussel, Dienst ELEM, B-1050 Brussels, Belgium. <sup>10</sup>Helmholtz-Zentrum Potsdam, Deutsches GeoForschungsZentrum GFZ, Department 1, Geodesy and Remote Sensing, Telegrafenberg A17, 14473 Potsdam, Germany. <sup>11</sup>Shell Technology Center, 560 048 Bangalore, India. <sup>12</sup>SRON Netherlands Institute for Space Research, PO Box 800, 9700 AV Groningen, The Netherlands. <sup>13</sup>Kapteyn Astronomical Institute, PO Box 800, 9700 AV Groningen, The Netherlands. <sup>14</sup>CSIRO Australia Telescope National Facility, PO Box 76, Epping, New South Wales 1710, Australia. <sup>15</sup>University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands. <sup>16</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. <sup>17</sup>Square Kilometre Array (SKA) South Africa, 3rd Floor, The Park, Park Road, Pinelands 7405, South Africa. <sup>18</sup>Institute for Astronomy, University of Edinburgh, Royal Observatory of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK. <sup>19</sup>University of Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany. <sup>20</sup>Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany. <sup>21</sup>School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK. <sup>22</sup>Research School of Astronomy and Astrophysics, Australian National University, Canberra, Australian Capital Territory 2611, Australia. <sup>23</sup>Anton Pannekoek Institute for Astronomy, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. <sup>24</sup>Max Planck Institute for Astrophysics, Karl Schwarzschild Strasse 1, 85741 Garching, Germany. <sup>25</sup>Onsala Space Observatory, Department of Earth and Space Sciences, Chalmers University of Technology, SE-43992 Onsala, Sweden. <sup>26</sup>SmarterVision BV, Oostersingel 5, 9401 JX Assen, The Netherlands. <sup>27</sup>Astronomisches Institut der Ruhr-Universität Bochum, Universitätsstrasse 150, 44780 Bochum, Germany. <sup>28</sup>Thüringer Landessternwarte, Sternwarte 5, D-07778 Tautenburg, Germany. <sup>29</sup>Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg. <sup>30</sup>Department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK. <sup>31</sup>Laboratoire Lagrange, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Boulevard de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France. <sup>32</sup>Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands. <sup>33</sup>LPC2E - Université d'Orléans/CNRS, 45071 Orléans Cedex 2, France. <sup>34</sup>Station de Radioastronomie de Nancy, Observatoire de Paris - CNRS/INSU, USR 704 - Université Orléans, OSUC, route de Souesmes, 18330 Nançay, France. <sup>35</sup>National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, New Mexico 87801-0387, USA. <sup>36</sup>Astro Space Center of the Lebedev Physical Institute, Profsoyuznaya street 84/32, Moscow 117997, Russia. <sup>37</sup>Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK. <sup>38</sup>National Astronomical Observatory of Japan, Tokyo 181-8588, Japan. <sup>39</sup>Sodankylä Geophysical Observatory, University of Oulu, Tähteläntie 62, 99600 Sodankylä, Finland. <sup>40</sup>STFC Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot OX11 0QX, UK. <sup>41</sup>Center for Information Technology (CIT), University of Groningen, PO Box 72, 9700 AB Groningen, The Netherlands. <sup>42</sup>Centre de Recherche Astrophysique de Lyon, Observatoire de Lyon, 9 avenue Charles André, 69561 Saint Genis Laval Cedex, France. <sup>43</sup>Fakultät für Physik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany. <sup>44</sup>Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa. <sup>45</sup>Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA. <sup>46</sup>GEPi, Observatoire de Paris, CNRS, Université Paris Diderot, 5 place Jules Janssen, 92190 Meudon, France. <sup>47</sup>LESIA, Observatoire de Paris, CNRS, UPMC, Université Paris Diderot, 5 place Jules Janssen, 92190 Meudon, France.



**Figure 1 | Energy resolution.** The distribution of  $f_r/f_p$  (blue bars) is fitted with a Gaussian (red dashed curve), yielding a standard deviation of  $\sigma = 0.12$  on a logarithmic scale, which corresponds to an energy resolution of 32%; this value is the quadratic sum of the energy resolution of the radio and particle resolutions. In this analysis, there was no absolute calibration for the received radio power, so  $f_r$  has an arbitrary scale.

telescope. Showers that occurred within an hour of lightning activity or that have a polarization pattern that is indicative of influences from atmospheric electric fields are excluded from the sample<sup>15</sup>.

Radio intensity patterns from air showers are asymmetric, owing to the interference between geomagnetic and charge-excess radiation. These patterns are reproduced from first principles by summing the radio contributions of all electrons and positrons in the shower. We use the radio simulation code CoREAS<sup>16</sup>, a plug-in of CORSIKA<sup>17</sup>, which follows this approach.

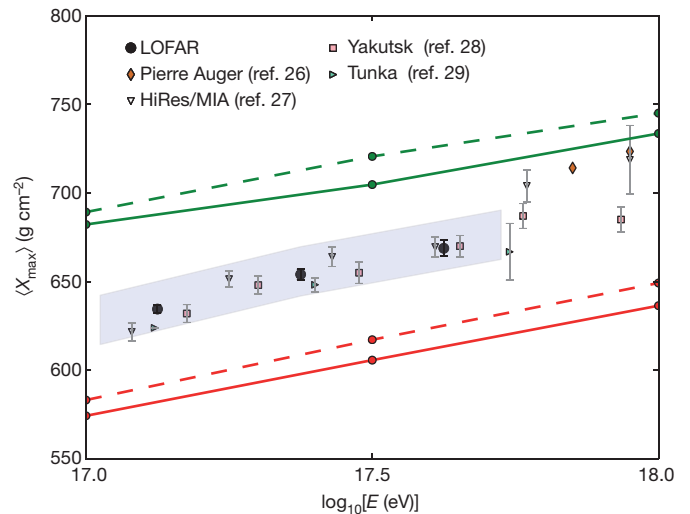
It has been shown that  $X_{\max}$ , the atmospheric depth of the shower maximum, can be accurately reconstructed from densely sampled radio measurements<sup>18</sup>. (The atmospheric depth is the air density integrated over the path that the particle has travelled, starting at the top of the atmosphere.) We use a hybrid approach that involves simultaneously fitting the radio and particle data. The radio component is very sensitive to  $X_{\max}$ , whereas the particle component is used for the energy measurement.

The fit contains four free parameters: the shower core position ( $x, y$ ), and scaling factors for the particle density ( $f_p$ ) and the radio power ( $f_r$ ). If  $f_p$  deviates substantially from unity, then the reconstructed energy does not match the simulation and a new set of simulations is produced. This procedure is repeated until the energies agree within the chosen uncertainties. The ratio of  $f_r$  and  $f_p$  should be the same for all showers, and is used to derive the energy resolution of 32% (see Fig. 1).

The radio intensity fits have reduced  $\chi^2$  values ranging from 0.9 to 2.9. All features in the data are well reproduced by the simulation (see Extended Data Figs 1–5), which demonstrates that the radiation mechanism is well understood. The reduced  $\chi^2$  values that exceed unity could indicate uncertainties in the antenna response or the atmospheric properties that were not already accounted for, or limitations of the simulation software.

Radio detection becomes more efficient for higher-altitude showers that have larger footprints (that is, larger areas on the ground in which the radio pulse can be detected). However, the particle trigger becomes less efficient because the number of particles reaching the ground decreases. To avoid a bias, we require that all the simulations produced for a shower satisfy a trigger criterion (see Methods). Above  $10^{17}$  eV, this requirement removes four showers from the sample. At lower energies, the number of showers excluded increases rapidly, and so we exclude all showers with energies less than  $10^{17}$  eV from our analysis.

Furthermore, we evaluate the reconstructed core positions of all simulated showers. Showers with a mean reconstruction error greater



**Figure 2 | Measurements of  $\langle X_{\max} \rangle$ .** Mean depth of the shower maximum  $X_{\max}$  as a function of energy  $E$  for LOFAR, and for previous experiments that used different techniques<sup>26–29</sup>. Error bars indicate  $1\sigma$  uncertainties. The systematic uncertainties are  $^{+14}_{-10} \text{ g cm}^{-2}$  on  $\langle X_{\max} \rangle$  and 27% on  $E$ , as indicated by the shaded band. The Pierre Auger Observatory<sup>26</sup> measures the fluorescent light emitted by atmospheric molecules excited by air-shower particles. HiRes/MIA<sup>27</sup> used a combination of this fluorescence technique and muon detection. The Yakutsk<sup>28</sup> and Tunka<sup>29</sup> arrays use non-imaging Cherenkov detectors. The green (upper) lines indicate  $\langle X_{\max} \rangle$  for proton showers simulated using QGSJETII.04 (solid) and EPOS-LHC (dashed); the red (lower) lines are for showers initiated by iron nuclei.

than 5 m are rejected. This criterion does not introduce a composition bias because it is based on the sets of simulated showers, not on the data. The final event sample contains 118 showers.

The uncertainty in  $X_{\max}$  is determined independently for all showers<sup>18</sup>, and has a mean value of  $16 \text{ g cm}^{-2}$  (see Extended Data Fig. 6). Figure 2 shows our measurements of the average  $X_{\max}$ ,  $\langle X_{\max} \rangle$ , which are consistent with earlier experiments using different methods. The high resolution for  $X_{\max}$  per shower allows us to derive more information about the composition of cosmic rays, by studying the shape of the  $X_{\max}$  distribution. For each shower, we calculate a mass-dependent parameter:

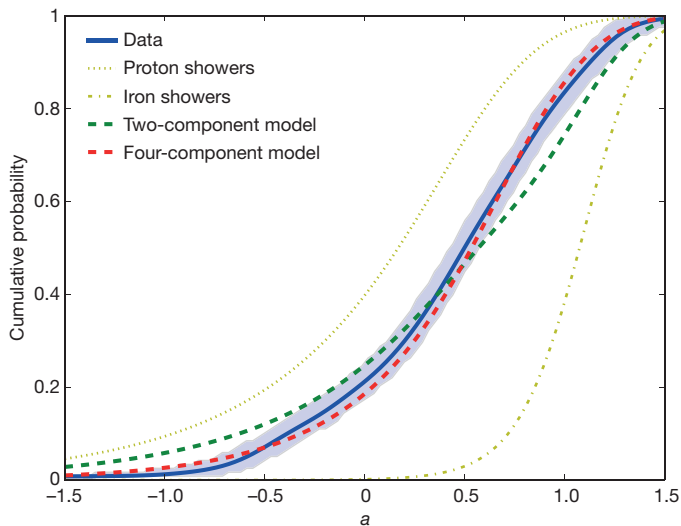
$$a = \frac{\langle X_{\text{proton}} \rangle - X_{\text{shower}}}{\langle X_{\text{proton}} \rangle - \langle X_{\text{iron}} \rangle} \quad (1)$$

in which  $X_{\text{shower}}$  is the reconstructed  $X_{\max}$ , and  $\langle X_{\text{proton}} \rangle$  and  $\langle X_{\text{iron}} \rangle$  are mean values of  $X_{\max}$  for proton and iron showers, respectively, predicted by the hadronic interaction code QGSJETII.04<sup>19</sup>.

The cumulative probability density function (CDF) for all showers is plotted in Fig. 3. First, we fit a two-component model of protons and iron nuclei (p and Fe), with the mixing ratio as the only free parameter. To calculate the corresponding CDFs we use a parameterization of the  $X_{\max}$  distribution fitted to simulations based on QGSJETII.04. The best fit is found for a proton fraction of 62%, but this fit describes the data poorly, with  $p = 1.1 \times 10^{-6}$ . (The test statistic for this fit is the maximum deviation between the data and the model CDFs, and  $p$  represents the probability of observing this deviation, or a larger one, assuming the fitted composition model; see Methods.)

A better fit is achieved with a four-component model of protons and helium, nitrogen and iron nuclei (p, He, N and Fe), yielding  $p = 0.17$ . Although the best fit is found for a helium fraction of 80%, the fit quality deteriorates slowly when replacing helium nuclei with protons. This is demonstrated in Fig. 4, in which  $p$  is plotted for four-component fits for which the fractions of helium nuclei and protons are fixed, and the ratio of nitrogen and iron nuclei is the only free parameter. The total fraction of light elements (p and He) is in the range [0.38, 0.98] at a 99% confidence level, with a best-fit value of 0.8. The heaviest





**Figure 3 | Composition model fits.** The cumulative probability density of the parameter  $a$  (see equation (1)) determined from the data (blue line; shading indicates the range in which  $p > 0.01$ ) and several models, on the basis of QGSJETII.04 simulations. The set that contains only proton showers is centred around  $a = 0$  and has a large spread (yellow dotted line), whereas iron showers give a distribution with a small spread centred around  $a = 1$  (yellow dash-dotted line). A two-component model (p and Fe; green dashed line) yields the best fit for a proton fraction of 62%, but does not describe the data well ( $p = 1.1 \times 10^{-6}$ ). A four-component model (p, He, N and Fe; red dashed line) gives the best fit with 0% protons, 79% helium, 19% nitrogen and 2% iron, with  $p = 0.17$ . The uncertainty on these values is presented in Fig. 4.

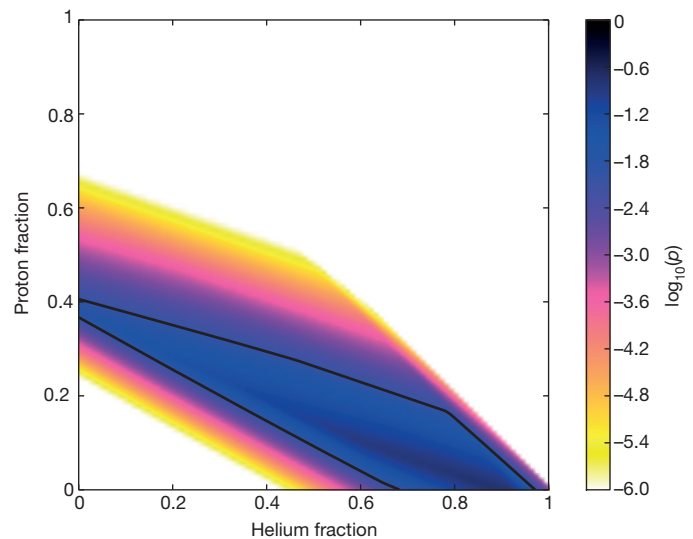
composition that is allowed within systematic uncertainties has a best-fit light-element fraction of 0.6 and a 99% confidence interval of [0.18, 0.82]. For information about the systematic uncertainties and the statistical analysis, see Methods.

The abundances of individual elements depend on the hadronic interaction model. The  $X_{\max}$  values predicted by EPOS-LHC<sup>20</sup> are, on average,  $15\text{--}20\text{ g cm}^{-2}$  higher than those predicted by QGSJETII.04 (see Fig. 2). This coincides with the separation in  $\langle X_{\max} \rangle$  between, for example, protons and deuterium or between helium and beryllium. Therefore, we present our result as a total fraction of light elements, to avoid placing too much emphasis on individual elements.

Recent results from the Pierre Auger Observatory<sup>3</sup> indicate that the composition of cosmic rays at  $10^{18}\text{ eV}$ , just below the ‘ankle’ (a hardening of the all-particle cosmic-ray spectrum), can be fitted with a mixture of protons and either helium (QGSJETII.04) or nitrogen (EPOS-LHC). As the energy decreases, the proton fraction of the cosmic-ray composition decreases while the helium (or nitrogen) fraction increases, down to the threshold energy of  $7 \times 10^{17}\text{ eV}$ . An extrapolation of this trend to our mean energy of  $3 \times 10^{17}\text{ eV}$  connects smoothly to our best-fitting solution in which helium dominates.

An ‘ankle’-like feature in the cosmic-ray energy spectrum at  $10^{17.1}\text{ eV}$  has been measured<sup>4</sup> at the KASCADE-Grande experiment, at which the spectral index for light elements changes to  $\gamma = -2.79 \pm 0.08$ . However, the light particle (p and He) fraction is found to be less than 30% at  $3 \times 10^{17}\text{ eV}$  (on the basis of figure 4 in ref. 4), which is considerably lower than our value. In contrast to LOFAR, the composition measurements presented in ref. 4 are based on the muon/electron ratio. A muon excess compared to all commonly used hadronic interaction models was reported<sup>21</sup>. Inaccurate predictions of muon production, or  $\langle X_{\max} \rangle$ , could be the cause of the discrepancy in the fraction of light particles predicted by LOFAR and KASCADE-Grande.

If the ‘knee’ in the all-particle cosmic-ray spectrum (a steepening near  $3 \times 10^{15}\text{ eV}$ ) corresponds to the proton or helium cut-off of the main galactic cosmic-ray population, then the corresponding iron cut-off would lie at an energy of at most 26 times larger. If the main



**Figure 4 |  $p$ -value distribution for the four-component model.** The four-component model is explored further by fixing the proton and helium fractions at all possible combinations, and solving for the nitrogen/iron ratio. The  $p$  value (see Fig. 3) is plotted as a function of the proton and helium fractions. The optimal fit (largest  $p$  value) is found for 0% protons and 79% helium ( $p = 0.17$ ), but the deviation deteriorates slowly when replacing helium with protons. The black contour line bounds all combinations for which  $p > 0.01$ . At this significance level, the total fraction of light elements (p and He) lies between 0.38 and 0.98.

population of galactic cosmic-ray sources still dominates at  $10^{17}\text{ eV}$ , then the mass composition of the cosmic rays should be dominated by heavy elements at that energy. Therefore, the large component of light elements observed with LOFAR must have another origin.

In principle, it is possible that we observe an extragalactic component. In that case, the ‘ankle’ in the cosmic-ray spectrum, at energies slightly greater than  $10^{18}\text{ eV}$ , does not indicate the transition from galactic to extragalactic origin. Instead, it can be explained as the imprint of pair production on the cosmic microwave background on an extragalactic proton spectrum<sup>22</sup>. However, because this feature only appears for a proton-dominated flux it is contrary to our data that indicate a mixture of light elements.

A second galactic component, dominating around  $10^{17}\text{ eV}$ , could be produced by a class of extremely energetic sources (galactic exatrons), such as the explosions of Wolf Rayet stars into their stellar winds<sup>23</sup> or past galactic gamma-ray bursts<sup>24</sup>. Alternatively, the original galactic population could be reaccelerated by the galactic-wind-termination shock<sup>25</sup>. Such scenarios predict mixtures of light elements, consistent with our results.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 19 March 2014; accepted 29 December 2015.**

1. IceCube Collaboration. Evidence for high-energy extraterrestrial neutrinos at the IceCube detector. *Science* **342**, 1242856 (2013).
2. Murase, K., Ahlers, M. & Lacki, B. Testing the hadronuclear origin of PeV neutrinos observed with IceCube. *Phys. Rev. D* **88**, 121301 (2013).
3. Aab, A. *et al.* Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications. *Phys. Rev. D* **90**, 122006 (2014).
4. Apel, W. *et al.* Ankle-like feature in the energy spectrum of light elements of cosmic rays observed with KASCADE-Grande. *Phys. Rev. D* **87**, 081101 (2013).
5. Kampert, K.-H. & Unger, M. Measurements of the cosmic ray composition with air shower experiments. *Astropart. Phys.* **35**, 660–678 (2012).
6. Allan, H. R. in *Progress in Elementary Particle and Cosmic Ray Physics* Vol. 10 (eds Wilson, J. G. & Wouthuysen, S. A.) 171–302 (North-Holland Pub. Co., 1971).
7. Falcke, H. & Gorham, P. W. Detecting radio emission from cosmic ray air showers and neutrinos with a digital radio telescope. *Astropart. Phys.* **19**, 477–494 (2003).
8. Falcke, H. *et al.* Detection and imaging of atmospheric radio flashes from cosmic ray air showers. *Nature* **435**, 313–316 (2005).

9. Huege, T. The renaissance of radio detection of cosmic rays. *Braz. J. Phys.* **44**, 520–529 (2014).
10. Apel, W. *et al.* Reconstruction of the energy and depth of maximum of cosmic-ray air showers from LOPES radio measurements. *Phys. Rev. D* **90**, 062001 (2014).
11. Belov, K. *et al.* Towards determining the energy of the UHECRs observed by the ANITA detector. *AIP Conf. Proc.* **1535**, 209–213 (2013).
12. Werner, K. & Scholten, O. Macroscopic treatment of radio emission from cosmic ray air showers based on shower simulations. *Astropart. Phys.* **29**, 393–411 (2008).
13. van Haarlem, M. *et al.* LOFAR: the LOw-Frequency ARray. *Astron. Astrophys.* **556**, A2 (2013).
14. Schellart, P. *et al.* Detecting cosmic rays with the LOFAR radio telescope. *Astron. Astrophys.* **560**, A98 (2013).
15. Schellart, P. *et al.* Probing atmospheric electric fields in thunderstorms through radio emission from cosmic-ray-induced air showers. *Phys. Rev. Lett.* **114**, 165001 (2015).
16. Huege, T., Ludwig, M. & James, C. Simulating radio emission from air showers with CoREAS. *AIP Conf. Proc.* **1535**, 128–132 (2013).
17. Heck, D., Knapp, J., Capdevielle, J. N., Schatz, G. & Thouw, T. *CORSIKA: a Monte Carlo code to simulate extensive air showers*. Report No. FZKA 6019 (Forschungszentrum Karlsruhe, 1998).
18. Buitink, S. *et al.* Method for high precision reconstruction of air shower  $X_{\text{max}}$  using two-dimensional radio intensity profiles. *Phys. Rev. D* **90**, 082003 (2014).
19. Ostapchenko, S. QGSJET-II: results for extensive air showers. *Nucl. Phys. B* **151**, 147–150 (2006).
20. Pierog, T. & Werner, K. EPOS model and ultra high energy cosmic rays. *Nucl. Phys. B* **196**, 102–105 (2009).
21. Aab, A. *et al.* Muons in air showers at the Pierre Auger Observatory: mean number in highly inclined events. *Phys. Rev. D* **91**, 032003 (2015).
22. Aloisio, R. *et al.* A dip in the UHECR spectrum and the transition from galactic to extragalactic cosmic rays. *Astropart. Phys.* **27**, 76–91 (2007).
23. Stanev, T., Biermann, P. & Gaisser, T. Cosmic rays. IV. The spectrum and chemical composition above 10 GeV. *Astron. Astrophys.* **274**, 902–915 (1993).
24. Calvez, A., Kusenko, S. & Nagataki, S. Role of galactic sources and magnetic fields in forming the observed energy-dependent composition of ultrahigh-energy cosmic rays. *Phys. Rev. Lett.* **105**, 091101 (2010).
25. Jokipii, J. R. & Morfill, G. Ultra-high-energy cosmic rays in a galactic wind and its termination shock. *Astrophys. J.* **312**, 170–177 (1987).
26. Letessier-Selvón, A. *et al.* Highlights from the Pierre Auger Observatory. *Braz. J. Phys.* **44**, 560–570 (2014).
27. Abu-Zayyad, T. *et al.* Measurement of the cosmic-ray energy spectrum and composition from  $10^{17}$  to  $10^{18.3}$  eV using a hybrid technique. *Astrophys. J.* **557**, 686–699 (2001).
28. Knurenko, S. & Sabourov, A. The depth of maximum shower development and its fluctuations: cosmic ray mass composition at  $E_0 \geq 10^{17}$  eV. *Astrophys. Space Sci. Trans.* **7**, 251–255 (2011).
29. Berezhnev, S. F. *et al.* Tunka-133: primary cosmic ray mass composition in the energy range  $6 \cdot 10^{15}$ – $10^{18}$  eV. *Proc. 32nd Int. Cosmic Ray Conf.* **1**, 209–212 (2011).

**Acknowledgements** We acknowledge financial support from the Netherlands Organization for Scientific Research (NWO), VENI grant 639-041-130, the Netherlands Research School for Astronomy (NOVA), the Samenwerkingsverband Noord-Nederland (SNN) and the Foundation for Fundamental Research on Matter (FOM). We acknowledge funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC (grant agreement no. 227610) and under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 640130). LOFAR, the Low Frequency Array designed and constructed by ASTRON, has facilities in several countries that are owned by various parties (each with their own funding sources) and that are collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy.

**Author Contributions** All authors are part of the LOFAR collaboration and have contributed to the design, construction, calibration and maintenance of LOFAR and/or LORA. The first thirteen authors constitute the Cosmic Ray Key Science Project and have contributed to the acquisition, calibration and analysis of cosmic-ray radio data and LORA data. The manuscript was written by S.B. and subjected to an internal collaboration-wide review process. All authors approved the final version of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.B. (Stijn.Buitink@vub.ac.be).

## METHODS

**Event selection.** Cosmic ray detection at LOFAR runs continuously in the background during astronomical observations. When 16 out of the 20 scintillator stations of the LORA particle array detect a signal, a ‘trigger’ is issued and the ring buffers of all active antennas within about a 1-km radius are stored for offline analysis<sup>14</sup>. Which antennas are active depends on the settings of the astronomical observation. For this analysis, we selected showers that were measured with at least four antenna stations (corresponding to at least 192 antennas) in the low band (30–80 MHz after filtering).

The trigger and selection criteria introduce a composition bias. This bias is removed by excluding certain showers on the basis of dedicated sets of simulations that are produced for each observed shower. Each of these sets contains 50 proton and 25 iron showers that span the whole range of possible shower depths. A shower is only accepted if all simulations in its set satisfy the trigger and selection criteria. This anti-bias exclusion removes many showers below  $10^{17}$  eV, but only four above that energy. Consequently, we restrict our analysis to the higher-energy showers, imposing a minimum bound on the reconstructed shower energy of  $E_{\text{reco}} = 10^{17}$  eV.

Imposing this energy bound introduces another potential source of compositional bias, because the reconstructed energy might depend on the depth of the shower. However, in our reconstruction approach, this effect is very small because energy and  $X_{\text{max}}$  are fitted simultaneously. Extended Data Fig. 7 shows distributions of the ratio between true and reconstructed energy for proton and iron simulations. The systematic offset between the two particle types is of the order of 1%.

We used data from the Royal Netherlands Meteorological Institute to check for lightning-storm conditions during our observations. When lightning strikes were detected in the north of the Netherlands within an hour from a detection, the event is flagged and excluded from the analysis. The presence of electric fields in the clouds can severely alter the radio emission even in the absence of lightning discharges<sup>30</sup>. The polarization angle of the radio pulse is very sensitive to the nature of the emission mechanism<sup>15,31</sup> and is used as an additional veto against strong field conditions.

Finally, a quality criterion is imposed on the sample so that only showers that have a core position and arrival direction that allows accurate reconstruction are included. We use the dedicated sets of simulations produced for each shower to derive uncertainties on core position, energy and  $X_{\text{max}}$ . These three values are highly correlated, so a single criterion based on the core uncertainty of  $\sigma_{\text{core}} < 5$  m is sufficient.

The quality criterion is based on the dedicated sets of simulations. These sets are produced for a specific combination of core position and arrival direction. Therefore, the quality criterion is effectively a criterion on position and direction, and does not introduce a composition bias.

There is no criterion on the quality of the reconstruction of the actual data. By applying the criteria described above we obtain a sample of 118 showers that are fitted to the simulation yielding reduced  $\chi^2$  values in the range 0.9–2.9. Deviations from unity can be ascribed to uncertainties in antenna response, atmospheric properties such as the index of refraction, or limitations of the simulation software.

**Reconstruction.** The energy and  $X_{\text{max}}$  of the shower are reconstructed with the technique described in ref. 18.

**Statistical uncertainty.** The statistical uncertainty on the power measurements of individual antennas includes three contributions. First, there is contribution from the background noise, which is a combination of system noise and the galactic background. Second, there is a contribution from uncertainties in the antenna response model. There can be differences between the responses of antennas, either because of antenna properties (for example, cross-talk between nearby antennas) or because of signal properties (for example, polarization). Because these fluctuations are different for each shower core position and arrival direction, they are essentially random and so are included as a 10% statistical uncertainty on the power. Third, there is a contribution due to the error introduced by interpolating the simulated pulse power. Strictly speaking this is not a measurement uncertainty, but it must be taken into account when fitting the data to simulation. The interpolation error is of the order of 2.5% of the maximum power<sup>18</sup>. The three contributions are added in quadrature and produce the  $1\sigma$  error bars shown in Extended Data Figs 1–5.

The statistical uncertainty on  $X_{\text{max}}$  is given by the quadratic sum of the uncertainties due to the reconstruction technique and the atmospheric correction. The former is found by applying our analysis to simulated events with added Gaussian noise, where the noise level is determined from the data.

In the CORSIKA simulations, the standard US atmosphere model was used. The reconstructed shower depth is corrected for variations in the atmosphere using data from the Global Data Assimilation System (GDAS) of the NOAA National Climatic Data Center. We follow a previously developed procedure<sup>32</sup>, which typically leads to adjustments of the order of 5–20 g cm<sup>−2</sup>. The remaining uncertainty after correction is of the order of 1 g cm<sup>−2</sup>.

The refractive index of air is a function of temperature, air pressure and relative humidity. Using local weather information, the final data were split in two groups of equal size, corresponding to conditions with relatively high or low refractive index. The mean reconstructed  $X_{\text{max}}$  of these two subsets deviate from that of the total sample by  $\pm 5$  g cm<sup>−2</sup>; we adopt this value as an additional statistical uncertainty. Because the refractivity used in simulation corresponds to dry air, there is also an associated systematic error (see below).

The total statistical uncertainty on  $X_{\text{max}}$  is found by adding the above factors in quadrature. A distribution of the uncertainty for the showers in our final sample is shown in Extended Data Fig. 6.

The energy resolution is 32% and is found by comparing energy scaling factors of the radio power and particle density fit (see Fig. 1).

**Systematic effects.** The data have been subjected to several tests (outlined below) to determine the systematic uncertainty on the reconstructed values for  $X_{\text{max}}$ .

**Zenith-angle dependence.** The final data are split into two groups of equal size by selecting showers with a zenith angle below or above 32°. For both groups, the mean reconstructed  $X_{\text{max}}$  is calculated, yielding deviations from the mean value of the complete sample of  $\pm 8$  g cm<sup>−2</sup>. This spread is larger than is expected from random fluctuations alone and is included as a systematic uncertainty. The dependence on zenith angle may be related to atmospheric uncertainties (see below).

**Refractive index of air.** As explained above, the refractive index changes because of differences in atmospheric conditions. Fluctuations in  $X_{\text{max}}$  due to changing humidity are of the order of 5 g cm<sup>−2</sup> with respect to the mean. However, the refractive index that was used in the radio simulations corresponds to dry air, and is a lower bound on the actual value. Therefore, the real value of  $X_{\text{max}}$  can be higher than the reconstructed value, but not lower; we adopt an asymmetric systematic uncertainty of  $+10$  g cm<sup>−2</sup>.

**Hadronic interaction model.** Because the reconstruction technique is based on full Monte Carlo simulations, it is sensitive to the choice of hadronic interaction model that is used. A comparison between QGSJETII.04, SYBILL 2.1 and EPOS-LHC, revealed that the uncertainty due to model dependence is about 5 g cm<sup>−2</sup>. The uncertainty on the composition due to different models (in other words, on how to interpret the measured  $X_{\text{max}}$  values) is larger.

**Radiation code.** For this analysis we used the radiation code CoREAS, in which the contributions of all individual charges to radiation field are added together. The advantage of this microscopic approach is that it is completely model-independent and based on first principles. ZHAireS<sup>33</sup> is another microscopic code, which gives very similar results<sup>34</sup>. To calculate the emission, CoREAS uses the end-point formalism<sup>35</sup>, whereas ZHAireS is based on the ZHS algorithm<sup>36</sup>. Both formalisms are derived directly from Maxwell's equations and have been shown to be equivalent<sup>37</sup>. The other difference between CoREAS and ZHAireS is that they take the particle distribution from different air-shower propagation codes (CORSIKA and AIRES, respectively) that internally use different hadronic interaction models. Because the radiation formalisms themselves are equivalent, small differences between CoREAS and ZHAireS are probably due to differences in the hadronic interaction models used to simulate the particle interactions. Therefore, the choice of radiation code does not introduce additional systematic uncertainty on top of the uncertainty due to hadronic interaction models that is already included. A comparison study with LOFAR data did not show any evidence for a systematic offset between the codes (S.B. *et al.*, in preparation).

The remaining small dependence of  $X_{\text{max}}$  on zenith angle is possibly related to the refractive index. Showers with different inclination angles have their shower maximum at different altitudes and, therefore, different local air pressures and refractive indices. Consequently, increasing the refractive index used in simulations will result in a zenith-dependent change in reconstructed  $X_{\text{max}}$ . This could potentially remove the observed dependence of the composition on zenith angle. Correctly taking into account a complete atmospheric model for the profile of the refractivity of air is subject of further study. Here, we treat the effect conservatively by linearly adding the first two contributions to the uncertainty. The other two contributions are independent and are added in quadrature, yielding a total systematic uncertainty of  $^{+14}_{-10}$  g cm<sup>−2</sup>.

The systematic uncertainty in the energy reconstruction with the LORA particle detector array is 27%, which includes effects due to detector calibration, hadronic interaction models and the assumed slope of the primary cosmic-ray spectrum in the CORSIKA simulations<sup>38,39</sup>.

**Statistical analysis.** For each observed shower, we calculate  $a$  using equation (1):

$$a = \frac{\langle X_{\text{proton}} \rangle - X_{\text{shower}}}{\langle X_{\text{proton}} \rangle - \langle X_{\text{iron}} \rangle}$$

in which  $X_{\text{shower}}$  is the reconstructed  $X_{\text{max}}$  and  $\langle X_{\text{proton}} \rangle$  and  $\langle X_{\text{iron}} \rangle$  are mean values predicted by QGSJETII.04<sup>19</sup>. Therefore,  $a$  is an energy-independent parameter that is mass sensitive. A pure proton composition results in a wide distribution of  $a$



centred around zero, whereas a pure iron composition would result in a narrower distribution centred around one.

From the measurements we construct a cumulative distribution function (CDF) using the following Monte Carlo approach. A realization of the data is made by taking the measured values for the energy and  $X_{\max}$ , adding random fluctuations based on the statistical uncertainty of these parameters, and calculating  $a$  and the corresponding CDF. By constructing a large number of realizations with different random fluctuations, we calculate the mean CDF and the region that contains 99% of all realizations. These are indicated in Fig. 3 as the solid blue line and the shaded region, respectively.

We fit theoretical CDFs on the basis of compositions with two or four mass components to the data. The test statistic in the fit is the maximum deviation between the data and the model CDFs. The  $p$  value represents the probability of observing this deviation, or a larger one, assuming the fitted composition model.

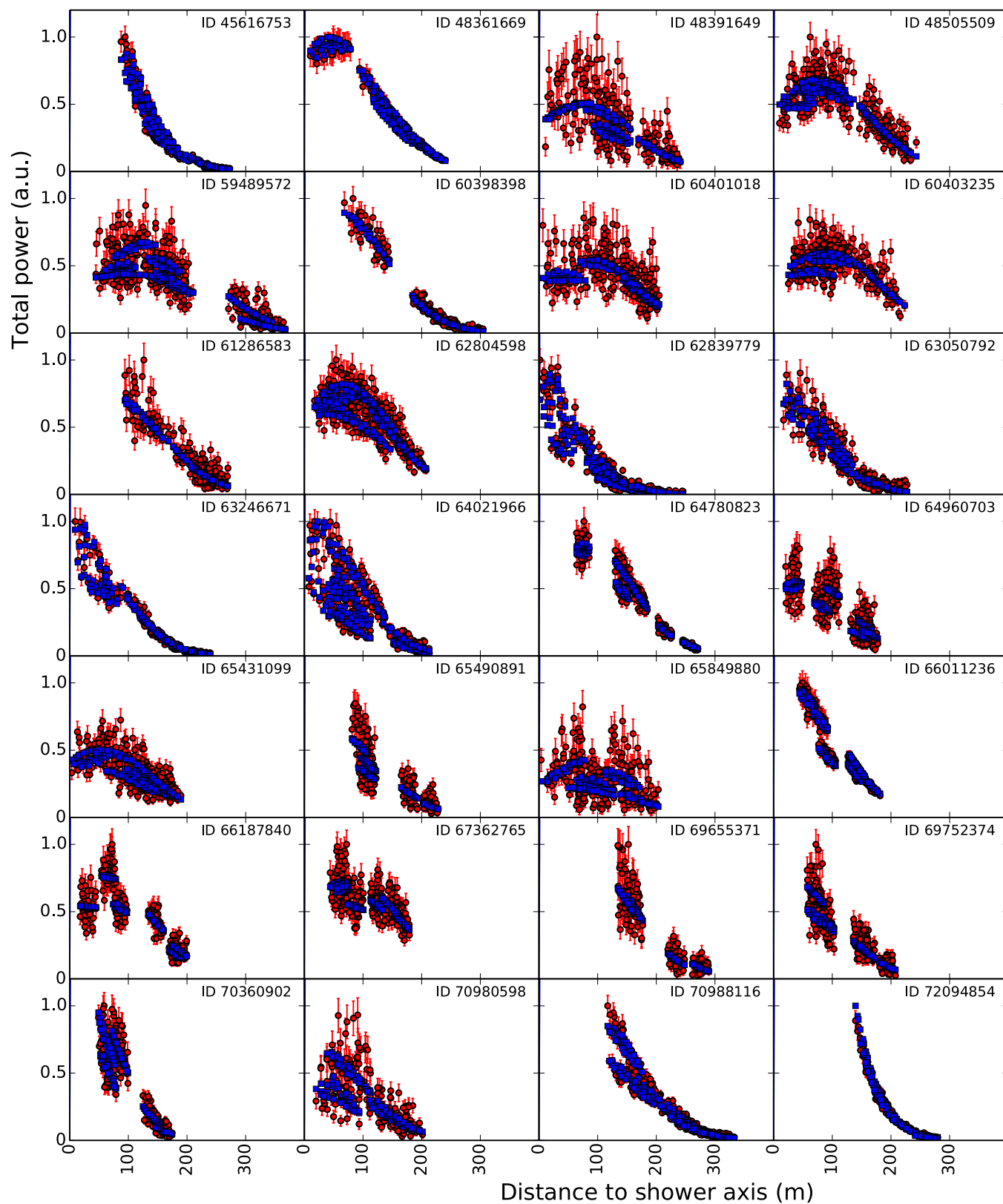
We first use a two-component model of proton and iron nuclei, in which the mixing ratio is the only free parameter. The best fit is found for a proton fraction of 62%, but it describes the data poorly, with  $p$  value of  $1.1 \times 10^{-6}$ .

A better fit is achieved with a four-component model (p, He, N and Fe), yielding  $p = 0.17$ . Although the best fit is found for a helium fraction of 80%, the fit quality deteriorates slowly when replacing helium by protons. This is demonstrated in Fig. 4, in which  $p$  is plotted for four-component fits with the fractions of helium and proton fixed, and the ratio between nitrogen and iron is the only free parameter. The solid line in Fig. 4 bounds the parameter space in which  $p > 0.01$ . We construct a 99% confidence interval on the total fraction of light elements (p and He) by finding the two extreme values of this fraction that lie within the  $p > 0.01$  region.

The total fraction of light elements (p and He) is in the range [0.38, 0.98] at the 99% confidence level, with a best fit value of 0.8. The heaviest composition that is allowed within systematic uncertainties (see above) has a best-fit p + He fraction of 0.6 and a 99% confidence interval of [0.18, 0.82].

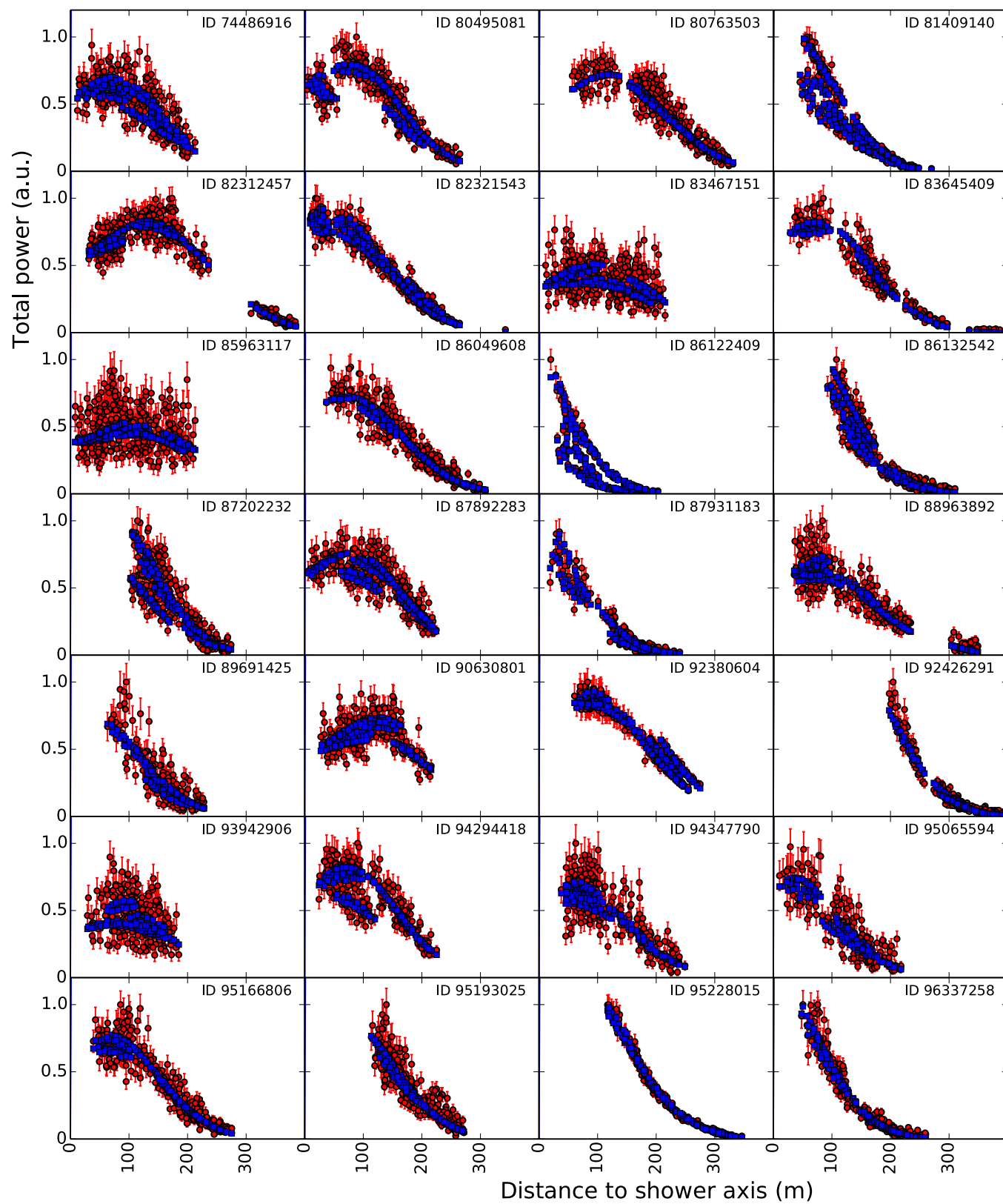
**Code availability.** Data analysis was done with PyCRTTools. PyCRTTools is free software, available from <http://usg.lofar.org/svn/code/trunk/src/PyCRTTools>, which can be redistributed and/or modified under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License or any later version.

30. Buitink, S. *et al.* Amplified radio emission from cosmic ray air showers in thunderstorms. *Astron. Astrophys.* **467**, 385–394 (2007).
31. Schellart, P. *et al.* Polarized radio emission from extensive air showers measured with LOFAR. *J. Cosmol. Astropart. Phys.* **10**, 014 (2014).
32. The Pierre Auger Collaboration. Description of atmospheric conditions at the Pierre Auger Observatory using the Global Data Assimilation System (GDAS). *Astropart. Phys.* **35**, 591–607 (2012).
33. Alvarez-Muñiz, J. *et al.* Monte Carlo simulations of radio pulses in atmospheric showers using ZHAireS. *Astropart. Phys.* **35**, 325–341 (2012).
34. Huege, T. *et al.* The convergence of EAS radio emission models and a detailed comparison of REAS3 and MGMR simulations. *Nucl. Instrum. Methods Phys. Res. A* **662** (Suppl.) S179–S186 (2012).
35. James, C., Falcke, H., Huege, T. & Ludwig, M. General description of electromagnetic radiation processes based on instantaneous charge acceleration in “endpoints”. *Phys. Rev. E* **84**, 056602 (2011).
36. Zas, E., Halzen, F. & Stanev, T. Electromagnetic pulses from high-energy showers: implications for neutrino detection. *Phys. Rev. D* **45**, 362–376 (1992).
37. Belov, K. *et al.* Accelerator measurements of magnetically induced radio emission from particle cascades with applications to cosmic-ray air showers. *Phys. Rev. Lett.* (in the press). Preprint at <http://arXiv.org/abs/1507.07296> (2015).
38. Thoudam, S. *et al.* LORA: a scintillator array for LOFAR to measure extensive air showers. *Nucl. Instrum. Methods Phys. Res. A* **767**, 339–346 (2014).
39. Thoudam, S. *et al.* Measurement of the cosmic-ray energy spectrum above  $10^{16}$  eV with the LOFAR Radboud Air Shower Array. *Astropart. Phys.* **73**, 34–43 (2016).



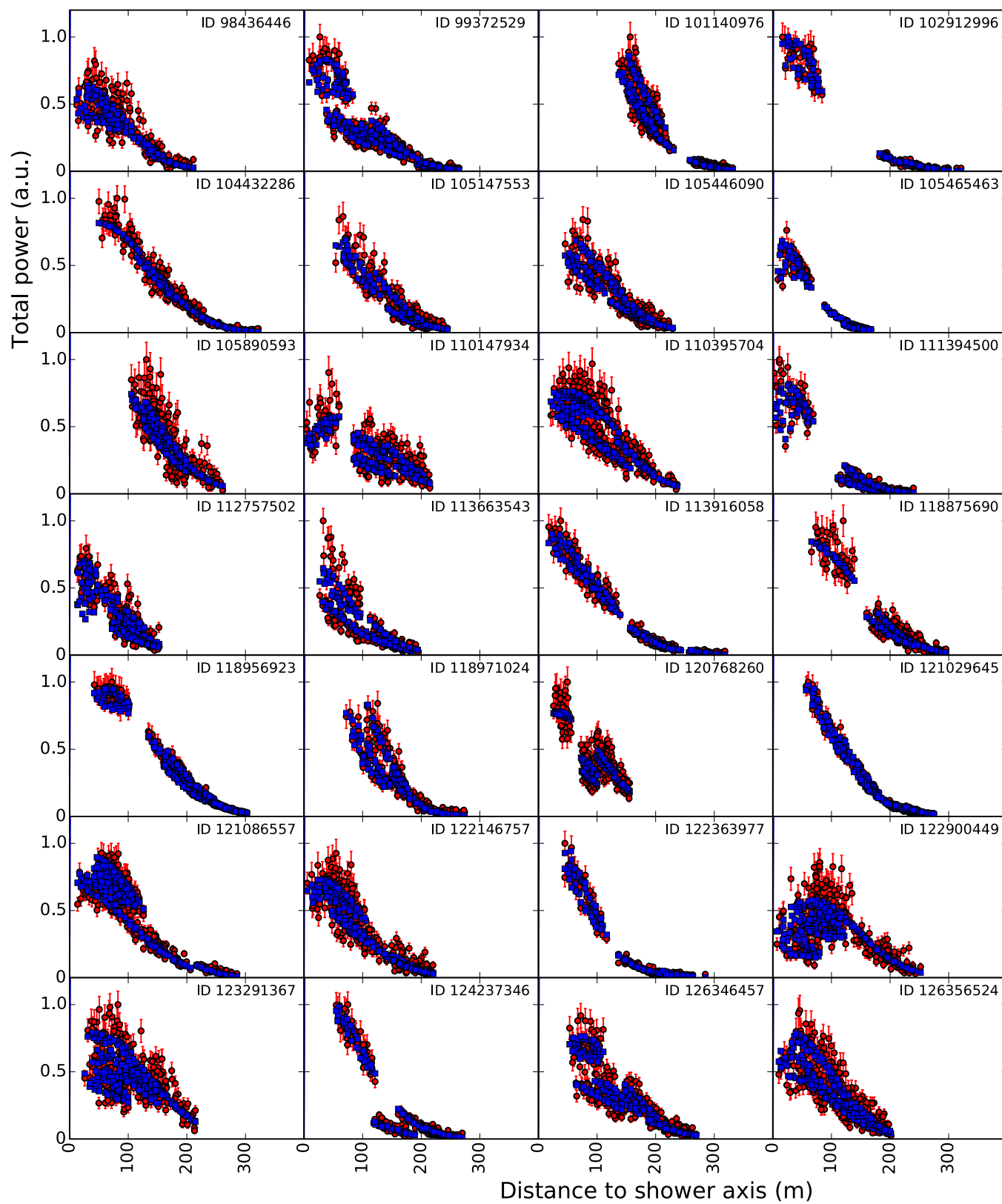
**Extended Data Figure 1 | Fitted lateral distributions.** Lateral distribution of radio-pulse power for all 118 measured showers (red circles) and the corresponding best-fitting CoREAS simulation (blue squares). The distance to the shower axis is the distance between the antenna and the

axis of the air shower. Therefore, a value of 0 corresponds to an antenna that is located at the position where the shower axis reaches the ground. The ID numbers are unique values that are used to label the detected air showers. a.u., arbitrary units.

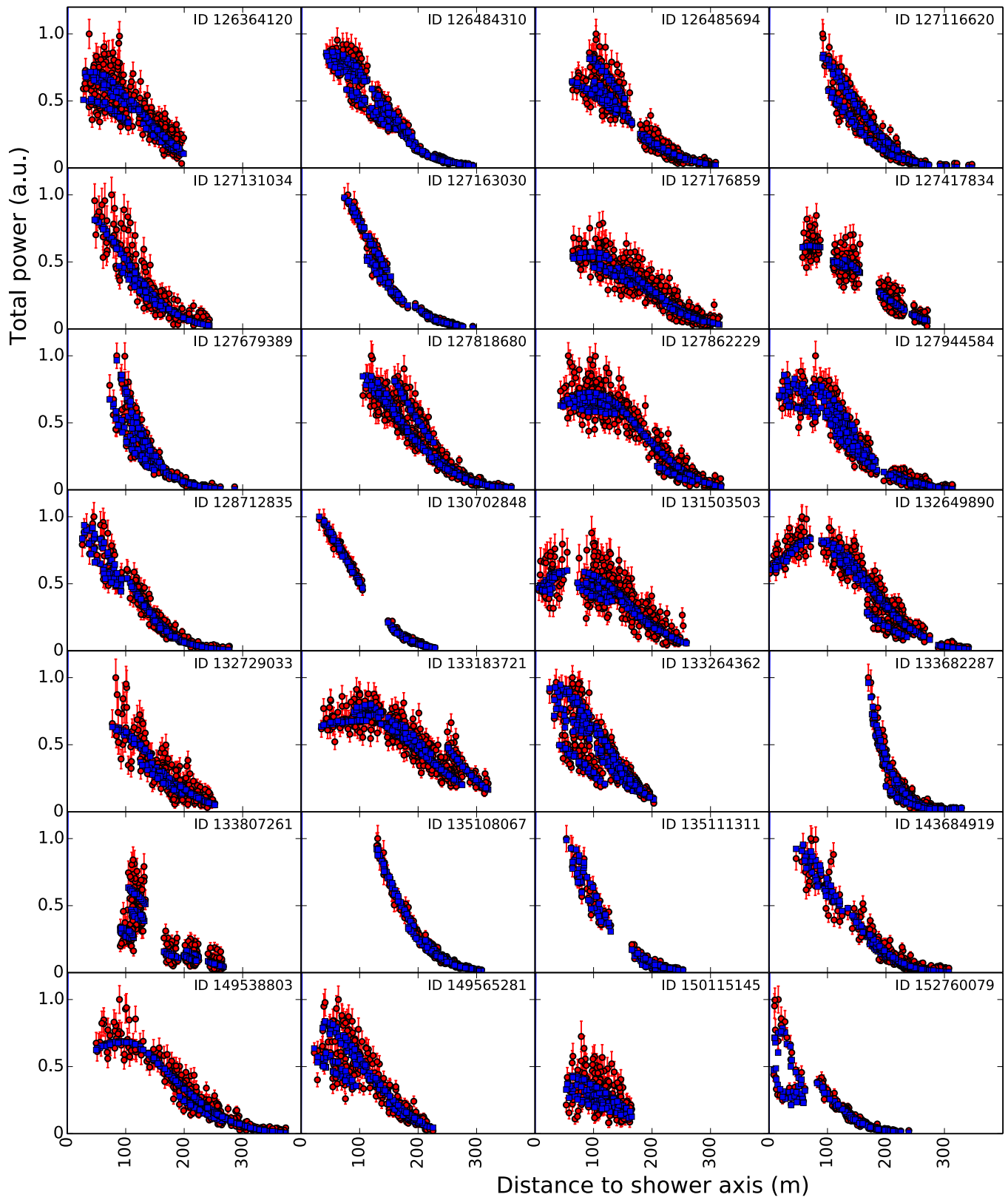


Extended Data Figure 2 | Fitted lateral distributions. Continuation of Extended Data Fig. 1.

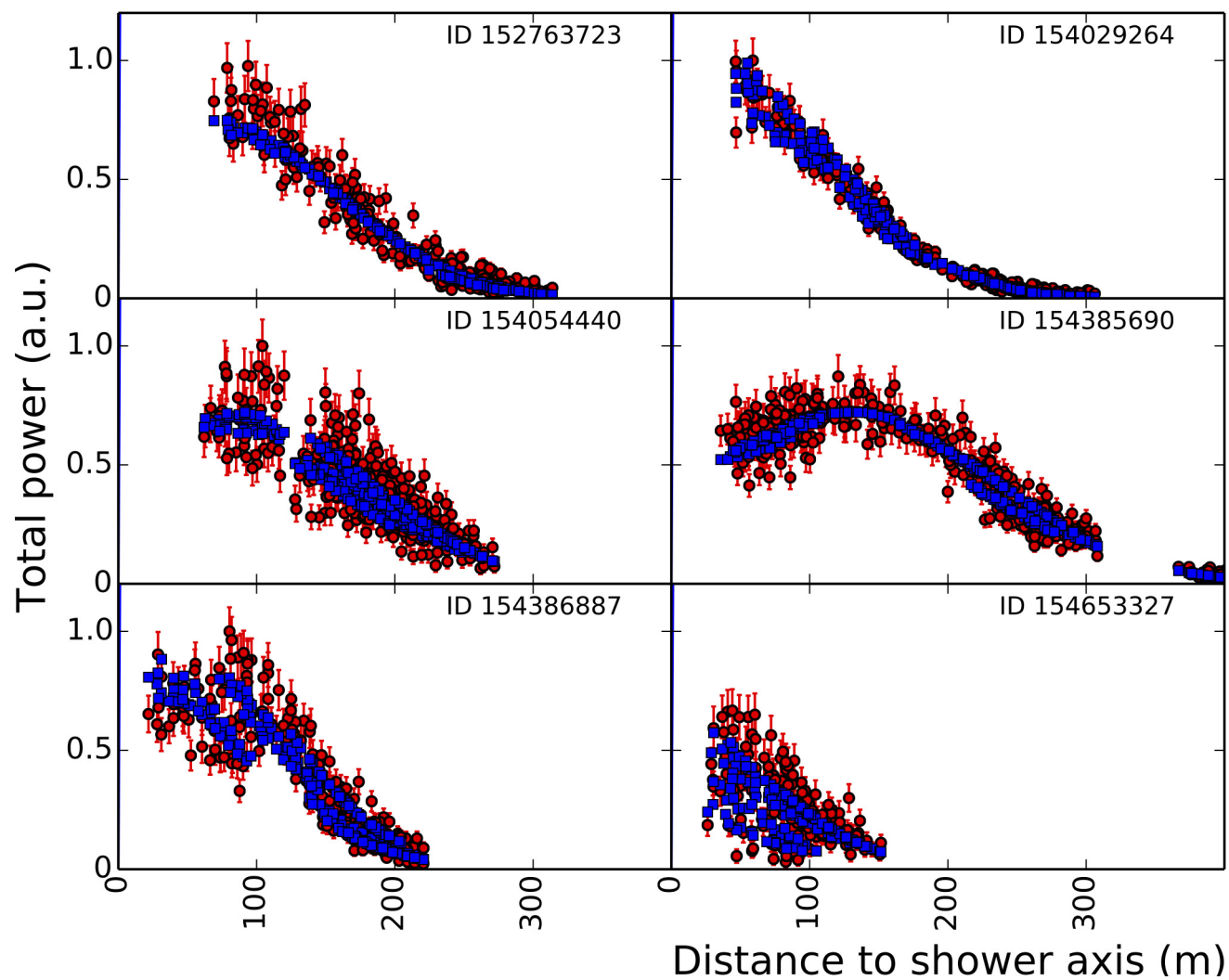




Extended Data Figure 3 | Fitted lateral distributions. Continuation of Extended Data Fig. 2.

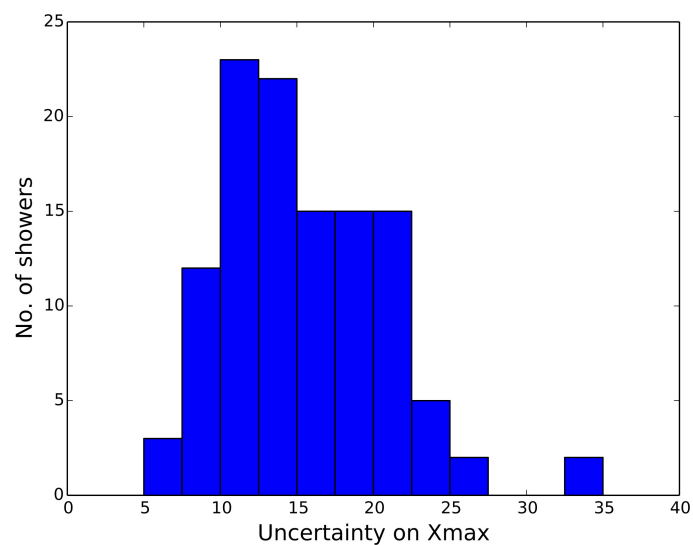


Extended Data Figure 4 | Fitted lateral distributions. Continuation of Extended Data Fig. 3.

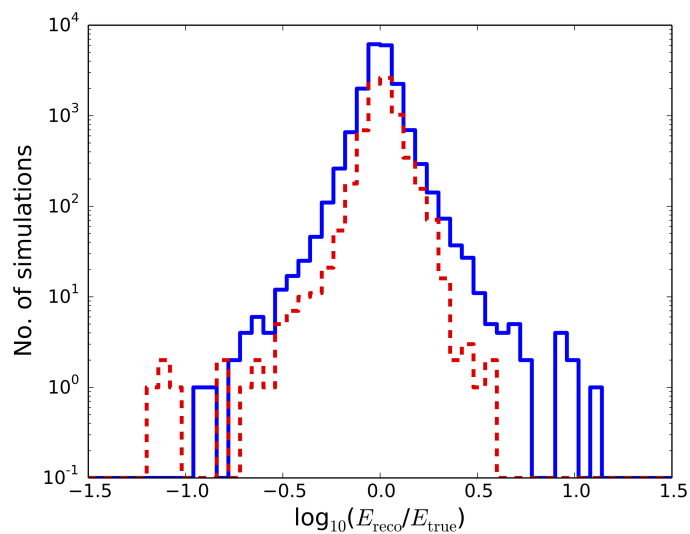


Extended Data Figure 5 | Fitted lateral distributions. Continuation of Extended Data Fig. 4.





**Extended Data Figure 6 | Distribution of uncertainty on  $X_{\text{max}}$ .** The distribution of the uncertainty on  $X_{\text{max}}$  for all showers used in this analysis. The mean value is  $16 \text{ g cm}^{-2}$ .



**Extended Data Figure 7 | Energy reconstruction.** Distributions of the ratio between true ( $E_{\text{true}}$ ) and reconstructed ( $E_{\text{reco}}$ ) energy for proton (blue solid line) and iron (red dashed line) showers. The two types of showers have a systematic offset of the order of 1%.

# Controlling spin relaxation with a cavity

A. Bienfait<sup>1</sup>, J. J. Pla<sup>2†</sup>, Y. Kubo<sup>1†</sup>, X. Zhou<sup>1,3</sup>, M. Stern<sup>1,4</sup>, C. C. Lo<sup>2</sup>, C. D. Weis<sup>5</sup>, T. Schenkel<sup>5</sup>, D. Vion<sup>1</sup>, D. Esteve<sup>1</sup>, J. J. L. Morton<sup>2</sup> & P. Bertet<sup>1</sup>

**Spontaneous emission of radiation is one of the fundamental mechanisms by which an excited quantum system returns to equilibrium. For spins, however, spontaneous emission is generally negligible compared to other non-radiative relaxation processes because of the weak coupling between the magnetic dipole and the electromagnetic field. In 1946, Purcell realized<sup>1</sup> that the rate of spontaneous emission can be greatly enhanced by placing the quantum system in a resonant cavity. This effect has since been used extensively to control the lifetime of atoms and semiconducting heterostructures coupled to microwave<sup>2</sup> or optical<sup>3,4</sup> cavities, and is essential for the realization of high-efficiency single-photon sources<sup>5</sup>. Here we report the application of this idea to spins in solids. By coupling donor spins in silicon to a superconducting microwave cavity with a high quality factor and a small mode volume, we reach the regime in which spontaneous emission constitutes the dominant mechanism of spin relaxation. The relaxation rate is increased by three orders of magnitude as the spins are tuned to the cavity resonance, demonstrating that energy relaxation can be controlled on demand. Our results provide a general way to initialize spin systems into their ground state and therefore have applications in magnetic resonance and quantum information processing<sup>6</sup>. They also demonstrate that the coupling between the magnetic dipole of a spin and the electromagnetic field can be enhanced up to the point at which quantum fluctuations have a marked effect on the spin dynamics; as such, they represent an important step towards the coherent magnetic coupling of individual spins to microwave photons.**

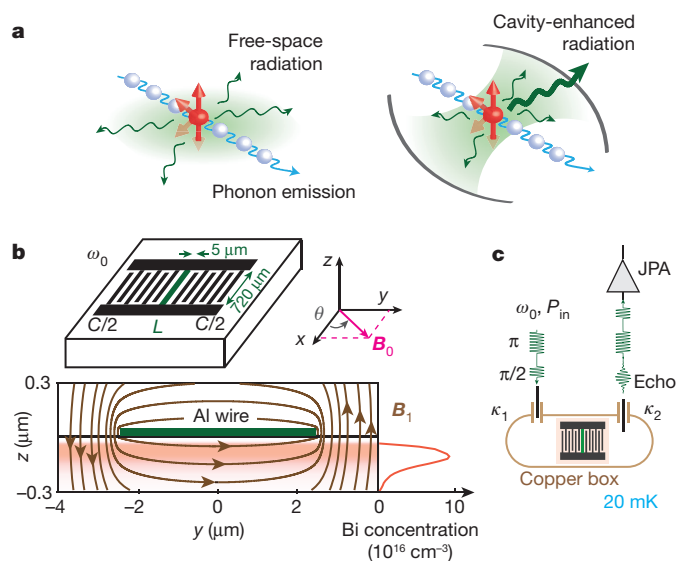
Spin relaxation is the process by which a spin reaches thermal equilibrium by exchanging an energy quantum  $\hbar\omega_s$  with its environment (where  $\hbar$  is the reduced Planck constant and  $\omega_s$  is the resonance frequency of the spin), for example in the form of a photon or a phonon, as shown in Fig. 1a. Understanding and controlling spin relaxation is essential in applications such as spintronics<sup>7</sup>, quantum information processing<sup>8</sup>, and magnetic resonance spectroscopy and imaging<sup>9</sup>. For such applications, the spin relaxation time  $T_1$  must be sufficiently long to permit coherent spin manipulation; however, if  $T_1$  is too long, it becomes a major bottleneck that limits the repetition rate of an experiment, which in turn affects factors such as the achievable sensitivity. Certain types of spins can be actively reset to their ground state by optical<sup>10</sup> or electrical<sup>11</sup> means, owing to their specific energy-level scheme, and methods such as chemical doping have been used to influence spin relaxation times *ex situ*<sup>12</sup>. Nevertheless, an efficient, general and tunable initialization method for spin systems is still currently lacking.

At first inspection, spontaneous emission would appear unlikely to influence spin relaxation: for example, an electron spin in free space and at a typical frequency of  $\omega_s/(2\pi) \approx 8$  GHz spontaneously emits photons at a rate of about  $10^{-12} \text{ s}^{-1}$ . However, the Purcell effect provides a way to markedly enhance spontaneous emission and thus gain precise and

versatile control over spin relaxation<sup>1</sup>. Consider a spin embedded in a microwave cavity of quality factor  $Q$  and frequency  $\omega_0$ . If the cavity damping rate  $\kappa = \omega_0/Q$  is greater than the spin–cavity coupling  $g$ , then the cavity provides an additional channel for spontaneous emission of microwave photons, governed by the Purcell rate<sup>6,13</sup>

$$\Gamma_P = \kappa \frac{g^2}{\kappa^2/4 + \delta^2} \quad (1)$$

in which  $\delta = \omega_0 - \omega_s$  is the spin–cavity detuning (see Fig. 1a and Methods). This cavity-enhanced spontaneous emission can be much larger than in free space, and is strongest when the spins and cavity are



**Figure 1 | Purcell-enhanced spin relaxation and experimental set-up.**

**a**, By placing a spin in a resonant cavity, radiative spin relaxation can be made to dominate over intrinsic processes such as phonon-induced relaxation. **b**, Top, a planar superconducting resonator with frequency  $\omega_0 = 1/\sqrt{LC}$  consisting of an interdigitated capacitor (black; with a capacitance  $C$ ) in parallel with an inductive wire (green; with an inductance  $L$ ) is fabricated on top of Bi-doped  $^{28}\text{Si}$ . A static magnetic field  $B_0$  is applied parallel to the  $x$ – $y$  plane of the 50-nm-thick aluminium layer, with a tunable orientation  $\theta$ . Bottom, magnetic field lines of the microwave excitation field  $B_1$  generated by the aluminium wire (arrows) are superimposed over the local concentration of Bi donors (red), obtained by secondary ion mass spectrometry (SIMS). **c**, The sample is mounted in a copper box that is thermally anchored at 20 mK, and probed by microwave pulses via asymmetric antennae that are coupled with rate  $\kappa_1 \approx \kappa_2/5$  to the resonator. Microwave pulses at  $\omega_0$  of power  $P_{\text{in}}$  are sent by antenna 1, and the microwave signal leaving via antenna 2 is directed to the input of a Josephson parametric amplifier (JPA).

<sup>1</sup>Quantronics Group, SPEC, CEA, CNRS, Université Paris-Saclay, CEA-Saclay, 91191 Gif-sur-Yvette, France. <sup>2</sup>London Centre for Nanotechnology, University College London, London WC1H 0AH, UK. <sup>3</sup>Institute of Electronics Microelectronics and Nanotechnology, CNRS UMR 8520, ISEN Department, Avenue Poincaré, CS 60069, 59652 Villeneuve d'Ascq Cedex, France. <sup>4</sup>Quantum Nanoelectronics Laboratory, BINA, Bar Ilan University, Ramat Gan, Israel. <sup>5</sup>Accelerator Technology and Applied Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>†</sup>Present addresses: School of Electrical Engineering & Telecommunications, University of New South Wales, Sydney, New South Wales 2052, Australia (J.J.P.); Okinawa Institute of Science and Technology (OIST) Graduate University, Onna, Okinawa 904-0495, Japan (Y.K.).



on resonance ( $\delta = 0$ ):  $\Gamma_p = 4g^2/\kappa$ . Furthermore, the Purcell rate can be modulated by changing the coupling constant or the detuning, allowing spin relaxation to be tuned on demand.

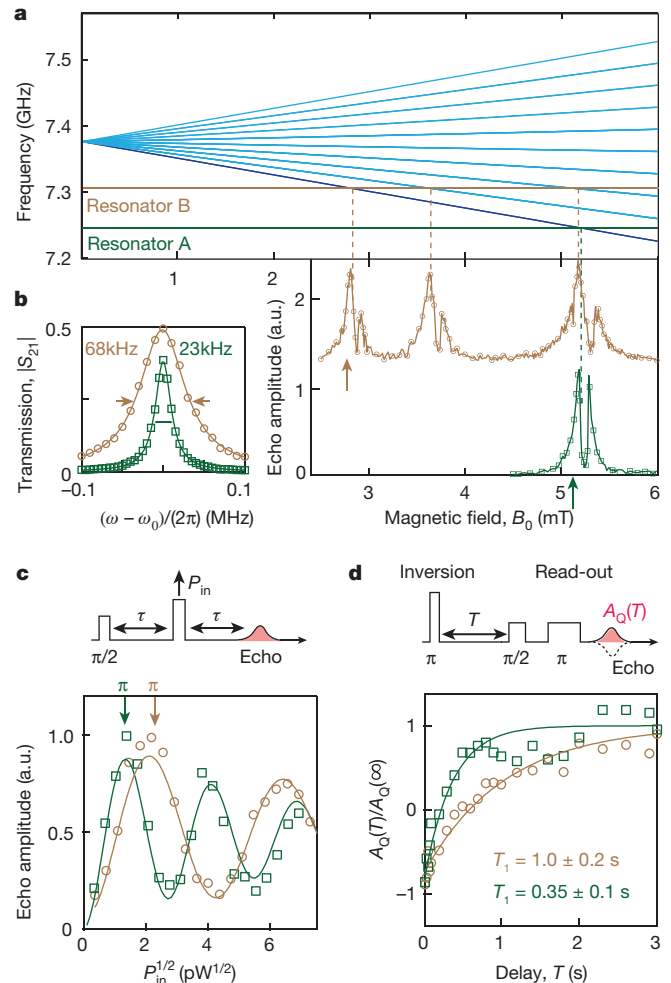
Although the Purcell effect was used to detect spontaneous emission of radiofrequency radiation from nuclear spins coupled to a resonant circuit<sup>14</sup>, the corresponding Purcell rate  $\Gamma_p \approx 10^{-16} \text{ s}^{-1}$  (or 1 photon emitted every 300 million years) was negligible compared to the intrinsic spin–lattice relaxation processes. For photon emission to become the dominant spin-relaxation mechanism, both a large spin–cavity coupling and a low cavity damping rate are needed; in our experiment, this is achieved by combining the microwave confinement provided by a micrometre-scale resonator with the high quality factors achieved by using superconducting circuits.

The device consists of two planar aluminium lumped-element superconducting resonators (denoted A and B) patterned onto a silicon chip that was purified in nuclear-spin-free  $^{28}\text{Si}$  and implanted with bismuth atoms (see Fig. 1b) at a sufficiently low concentration for collective radiation effects to be absent. A static magnetic field  $B_0$  is applied in the plane of the aluminium resonators, at an angle  $\theta$  from the resonator inductive wire, tunable *in situ*. The device is mounted inside a copper box and cooled to 20 mK. Each resonator can be used to perform inductive detection of the electron-spin resonance (ESR) signal of the bismuth donors: microwave pulses at  $\omega_0$  are applied at the resonator input, generating an oscillating magnetic field  $B_1$  around the inductive wire that drives the surrounding spins; the quantum fluctuations of this field, present even when no microwave is applied, are responsible for the Purcell spontaneous emission. Hahn echo pulse sequences<sup>15</sup> are used, resulting in the emission of a spin-echo in the detection waveguide, which is amplified with a sensitivity reaching the quantum limit by a Josephson parametric amplifier<sup>16</sup> before demodulation at room-temperature, yielding the integrated echo signal quadrature  $A_Q$  (see Methods). A more detailed description of the set-up is found in ref. 17.

Bismuth is a donor in silicon<sup>18</sup> with a nuclear spin  $I = 9/2$ . At cryogenic temperatures it can bind an electron (with spin  $S = 1/2$ ) in addition to those shared with the surrounding Si lattice. The large hyperfine interaction  $AS \cdot I$  between the electron and nuclear spin (in which  $S$  and  $I$  are the electron and nuclear spin operators, and  $A/h = 1.475 \text{ GHz}$  with  $h$  the Planck constant) produces a splitting of 7.375 GHz between the ground and excited multiplets at zero magnetic field (see Fig. 2a for the complete energy diagram<sup>19</sup>). This splitting makes the system ideal for coupling to superconducting circuits<sup>20,21</sup>. At low fields ( $B_0 < 10 \text{ mT}$ , compatible with the critical field of aluminium), all  $\Delta m_F = \pm 1$  transitions are allowed, where  $m_F$  is the projection of the total spin ( $F = I + S$ ) along  $B_0$ . Considering only the transitions with largest matrix element, resonator A ( $\omega_{0,A}/(2\pi) = 7.245 \text{ GHz}$ ,  $Q_A = 3.2 \times 10^5$ ) crosses the  $|F, m_F\rangle = |4, -4\rangle \leftrightarrow |5, -5\rangle$  transition, whereas resonator B ( $\omega_{0,B}/(2\pi) = 7.305 \text{ GHz}$ ,  $Q_B = 1.1 \times 10^5$ ) crosses the transitions  $|4, -4\rangle \leftrightarrow |5, -5\rangle$ ,  $|4, -3\rangle \leftrightarrow |5, -4\rangle$  and  $|4, -2\rangle \leftrightarrow |5, -3\rangle$  (see Fig. 2a, b).

The echo signal  $A_Q$  from each resonator as a function of  $B_0$  shows resonances at the expected magnetic fields, split into two peaks each with a full-width at half-maximum of  $\Delta\omega/(2\pi) \approx 2 \text{ MHz}$  (see Fig. 2a). As is explained in ref. 17, this splitting is believed to be the result of strain induced in the silicon at the donor implant depth of approximately 100 nm by the aluminium circuit deposited on the surface. In the following, we focus on the lower-frequency peak of the  $|4, -4\rangle \leftrightarrow |5, -5\rangle$  line, which corresponds to spins lying under the wire. Over the region occupied by these spins, the amplitude of the  $B_1$  field varies by less than  $\pm 2\%$ , as evidenced by the well-defined Rabi oscillations observed when we sweep the power of the refocusing pulse  $P_{\text{in}}$  at the cavity input (see Fig. 2c), which allows us to determine the input power of a  $\pi$  pulse for a given pulse duration.

We measure the relaxation time  $T_1$  by performing an “inversion-recovery” experiment<sup>22</sup> (see schematic in Fig. 2d), with the static field  $B_0$  aligned along  $x$  ( $\theta = 0$ ). A  $\pi$  pulse first inverts the spins whose frequencies lie within the resonator bandwidth  $\kappa_A/(2\pi) = 23 \text{ kHz}$  or



**Figure 2 | ESR spectroscopy and Purcell-limited  $T_1$  measurement.**

**a**, Top, dominant electron spin resonance transitions of the  $\text{Si}^{209}\text{Bi}$  spin system (see Methods). We use two resonators, A (green) and B (brown), with frequencies of 7.246 GHz and 7.305 GHz, respectively, that cross up to three spin transitions in the magnetic field range 0–6 mT, as seen in the echo-detected magnetic field sweep (bottom; vertically offset for clarity). Subsequent spin relaxation measurements were made at the magnetic fields indicated by the arrows, corresponding to the  $|F, m_F\rangle = |4, -4\rangle \leftrightarrow |5, -5\rangle$  transition for each resonator. The doublet structure of each transition is caused by strain exerted by the aluminium film on the donors<sup>17</sup>. **b**, Cavity linewidths for resonators A and B are found to be 23 kHz and 68 kHz, respectively, from fits (solid lines) to their measured transmission amplitude. **c**, Rabi oscillations are driven by varying the cavity input power of the refocusing  $\pi$  pulse (5  $\mu\text{s}$  long) applied  $\tau = 300 \mu\text{s}$  after the first  $\pi/2$  pulse. Solid lines are exponentially damped sinusoidal fits. **d**, The inversion-recovery sequence is used to measure the spin relaxation time  $T_1$ . Spin polarization is measured with a Hahn echo sequence.  $A_Q$  is rescaled by its value for  $T \gg T_1$  ( $A_Q(\infty)$ ) such that it varies from  $-1$  when the spins are fully inverted to  $+1$  at thermal equilibrium (see Methods for full sequence description). Data were obtained with the static field  $B_0$  parallel to the inductor ( $\theta = 0$ ). Solid lines are exponential fits to the data with time constant  $T_1$ . The uncertainty is provided by the standard deviation in the exponential fit parameters. a.u., arbitrary units. In all panels, the symbols represent data for each resonator (A, green squares; B, brown circles).

$\kappa_B/(2\pi) = 68 \text{ kHz}$ ; this constitutes a small subset of the total number of spins because  $\kappa_{A,B} \ll \Delta\omega$ . After a varying delay  $T$ , a Hahn echo sequence provides a measure of the longitudinal spin polarization. By fitting the data with decaying exponentials, we extract  $T_1 = 0.35 \text{ s}$  for resonator A and  $T_1 = 1.0 \text{ s}$  for resonator B.

To quantitatively compare our results with the expected Purcell rate, it is necessary to evaluate the spin–resonator coupling constant

$g = \gamma_e \langle F, m_F | S_x | F+1, m_F-1 \rangle \| \delta \mathbf{B}_\perp \|$ , in which  $\gamma_e/(2\pi) \approx 28 \text{ GHz T}^{-1}$  is the electronic gyromagnetic ratio,  $S_x$  is the dimensionless Pauli operator for the electron spin and  $\delta \mathbf{B}_\perp$  is the component of the resonator-field vacuum fluctuations orthogonal to  $\mathbf{B}_0$  (see Methods). A numerical estimate yields  $g_0/(2\pi) = 56 \pm 1 \text{ Hz}$  for the spins located below the inductive wire in the resonator that are probed in our measurements, and for  $\theta = 0$ . An independent estimate is obtained by measuring Rabi oscillations: their frequency  $\Omega_R = 2g_0\sqrt{n}$  directly yields  $g_0$  given knowledge of the average intra-cavity photon number  $\bar{n}$ , which can be determined with about 30% imprecision from  $P_{\text{in}}$  and the measured resonator coupling to the input and output antennae (see Methods). Using this method, we obtain  $g_0/(2\pi) = 50 \pm 7 \text{ Hz}$  for resonator A and  $58 \pm 7 \text{ Hz}$  for resonator B, compatible with the numerical estimate. The corresponding Purcell time of the resonant spontaneous emission is  $\Gamma_p^{-1} = 0.36 \pm 0.09 \text{ s}$  for resonator A and  $\Gamma_p^{-1} = 0.81 \pm 0.17 \text{ s}$  for resonator B, in agreement with the experimental values.

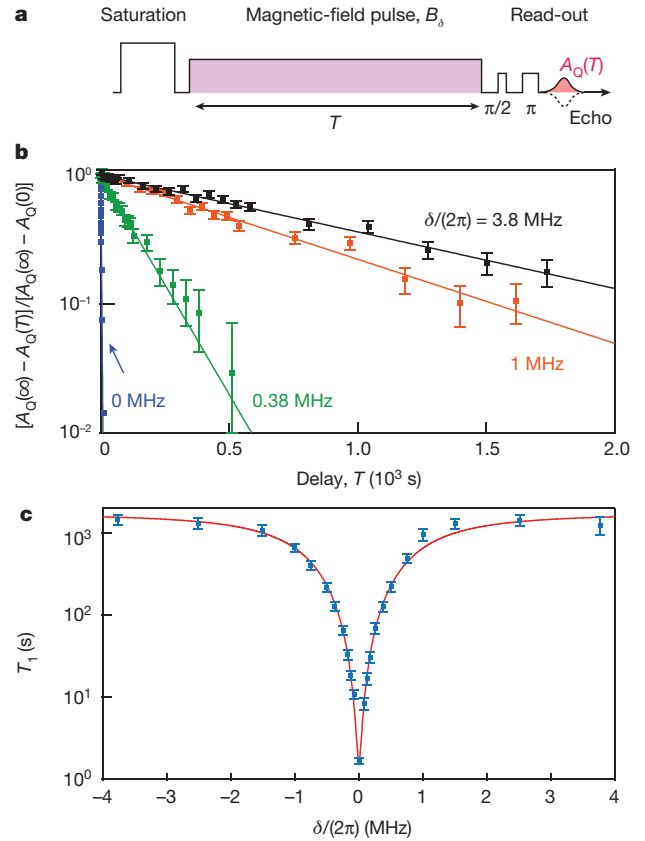
According to equation (1), a Purcell-limited  $T_1$  should be strongly dependent on the spin–cavity detuning. We introduce a pulse in the magnetic field of duration  $T$  between the spin excitation and the spin-echo sequence (see Fig. 3a), which results in a temporary detuning  $\delta$  of the spins. The amplitude of the echo signal  $A_Q$  as a function of  $T$  yields their energy relaxation time while they are detuned by  $\delta$ . To minimize the influence of spin diffusion<sup>22</sup>, the spin excitation is performed by a high-power long-duration saturating pulse (see Fig. 3a and Methods) instead of an inversion pulse as in Fig. 2d. As is evident in Fig. 3b, we find that the decay of the echo signal is well fitted by a single exponential with a decay time that increases with  $|\delta|$ . The extracted  $T_1(\delta)$  curve (see Fig. 3c) shows an increase in  $T_1$  of up to three orders of magnitude when the spins are detuned away from resonance, until it becomes limited by a non-radiative energy decay mechanism with characteristic time  $\Gamma_{\text{NR}}^{-1} = 1,600 \pm 300 \text{ s}$ . Given the doping concentration in our sample, this non-radiative decay time is consistent with earlier measurements of donor spin relaxation times<sup>23</sup>, which have been attributed to charge hopping, but it could also arise here from spatial diffusion of the spin magnetization away from the resonator mode volume. It is shown in Fig. 3c that the  $T_1(\delta)$  measurements are in agreement with the expected dependence  $(\Gamma_p(\delta) + \Gamma_{\text{NR}})^{-1}$ , with  $\Gamma_{\text{NR}}$  the only free parameter in this fit.

Having demonstrated the effect of cavity linewidth and detuning on the Purcell rate, we explore the effect of modulating the spin–cavity coupling constant  $g$ . This can be achieved by varying the orientation  $\theta$  of the static magnetic field  $\mathbf{B}_0$  in the  $x$ – $y$  plane (Fig. 1b), which adjusts the component of the microwave magnetic field ( $\mathbf{B}_1$ , which is mostly aligned along  $y$  under the inductive wire) that is orthogonal to  $\mathbf{B}_0$ . More precisely

$$g(\theta) = \gamma_e \langle F, m_F | S_x | F+1, m_F-1 \rangle \sqrt{\delta B_{1,y}^2 \cos^2(\theta) + \delta B_{1,z}^2} \quad (2)$$

(since  $\delta B_{1,x} = 0$ ). This orientation dependence is verified experimentally by measuring the Rabi frequency as a function of  $\theta$ , as shown in Fig. 4a, b, which allows us to extract  $g(0)/(2\pi) = 58 \text{ Hz}$  and  $g(\pi/2)/(2\pi) = 17 \text{ Hz}$ . As expected, we measure longer spin relaxation times for increasing values of  $\theta$ , as shown in Fig. 4c, with the relaxation rate  $T_1^{-1}$  scaling as  $[g(\theta)]^2$ , in agreement with equation (1). Overall, the data in Figs 3 and 4 demonstrate unambiguously that cavity-enhanced spontaneous emission is by far the dominant spin-relaxation channel when the spins are resonant with the cavity, because the probability of a spin-flip occurring as a result of emission of a microwave photon in the cavity is  $1/[1 + \Gamma_{\text{NR}}/\Gamma_p(\delta=0)] = 0.999$ , very close to unity.

The spontaneous emission evidenced here is an energy-relaxation mechanism that does not require the presence of a macroscopic magnetization to be effective. Under the Purcell effect, each spin independently relaxes towards thermal equilibrium by microwave photon emission, so that when no intra-cavity thermal field is present, the sample ends up in a fully polarized state after a time longer than  $\Gamma_p^{-1}$ , regardless of its initial state. This is in stark contrast to the well-known

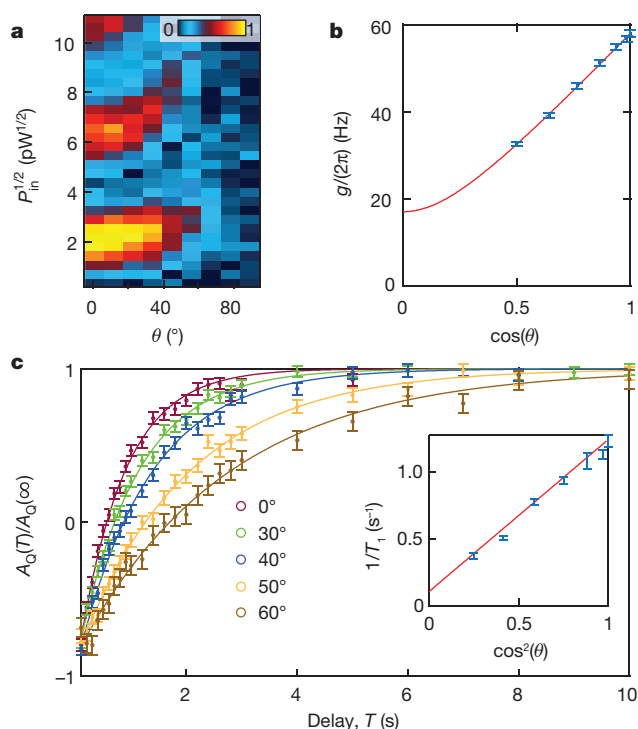


**Figure 3 | Controlling Purcell relaxation by spin–cavity detuning.**

**a**, In between their saturation and subsequent read-out, the spins are detuned from the cavity by  $\delta = \frac{d\omega_s}{dB} B_\delta$  by applying a magnetic-field pulse with an amplitude of  $B_\delta$ , with  $\frac{1}{2\pi} \frac{d\omega_s}{dB} \approx 25 \text{ GHz T}^{-1}$  for this transition and magnetic field. **b**, Measured spin-polarization decays (symbols) for four different detunings  $\delta$ , which are well fitted by exponential decays (lines), with relaxation time constants  $T_1$  increasing with the detuning (error bars indicate the standard deviation of a measured echo). **c**, Measured  $T_1$  as a function of detuning  $\delta$  (blue symbols). The red line is a fit with  $(\Gamma_p(\delta) + \Gamma_{\text{NR}})^{-1}$ , yielding  $\Gamma_{\text{NR}}^{-1} = 1,600 \text{ s}$ . Error bars are estimates of the standard deviation of the fit. These measurements are taken using resonator B and with  $\theta = \pi/4$ , which results in  $T_1 = 1.7 \text{ s}$  at  $\delta = 0$ .

phenomenon of radiative damping<sup>24</sup> of a transverse magnetization generated by earlier microwave pulses, which is a coherent collective effect under which the degree of polarization of a sample cannot increase. Had our device possessed a larger spin concentration, spontaneous relaxation would have occurred collectively, manifesting itself as a non-exponential decay of the echo signal on a timescale faster than  $\Gamma_p^{-1}$  (ref. 13), and leading to an incomplete thermalization<sup>6,25</sup>. The existence of such super-radiant or maser emission<sup>26</sup> requires the dimensionless ‘co-operativity’ parameter  $C = Ng^2/(\kappa\Delta\omega)$  (where  $N$  is the total number of spins) to satisfy  $C \gg 1$  (refs 6, 25, 27), which is not the case here because of the large inhomogeneous broadening of the spin resonance caused by strain.

Our demonstrated ability to modulate spin relaxation through three orders of magnitude by changing the applied field by less than 0.1 mT opens up new perspectives for spin-based quantum information processing: long intrinsic relaxation times, which are desirable to maximize the spin coherence time, can be combined with fast, on-demand initialization of the spin state. Similarly, performing electron spin resonance at dilution refrigerator temperatures can be prohibitively slow without the ability to accelerate spin relaxation on demand. We also anticipate that Purcell relaxation will offer a powerful approach to dynamical nuclear polarization<sup>28,29</sup>, for example, by tuning the cavity to match an electron–nuclear spin flip–flop transition, enhancing the rate of



**Figure 4 | Dependence of Purcell relaxation on spin-cavity coupling  $g$ .** **a**, Rabi oscillations (as in Fig. 2c) measured as a function of field orientation  $\theta$  (see Fig. 1b); the colour scale indicates the echo amplitude in arbitrary units. **b**, The Rabi oscillations in **a** are used to extract the spin-cavity coupling strength  $g$  (blue symbols; error bars are determined by the 30% accuracy on  $P_{\text{in}}$ ). These data are fit to equation (2) (red line); the non-zero value of  $g(\pi/2)$  is due to the finite out-of-plane component of the microwave magnetic field. **c**, Inversion-recovery measurements (error bars indicate the standard deviation of a measured echo) for different values of  $\theta$  confirm that the relaxation time  $T_1$  (see inset; error bars are estimates of the standard deviation of the fit) varies as  $[g(\theta)]^2$ . The red line in the inset is the Purcell formula predicted using the  $g(\theta)$  dependence fitted from **b**. All data were collected using resonator B.

cross-relaxation to pump polarization into the desired nuclear spin state<sup>30</sup> (see Methods). The Purcell rate we obtain could be increased by reducing the transverse dimensions of the inductor wire to yield larger coupling constants (up to 5–10 kHz), which would reduce the spontaneous emission time to less than 1 ms (thus permitting faster repetition rates and a higher sensitivity<sup>17</sup>), allowing for the possibility of high-co-operativity coupling of a single spin to the microwave cavity field. Our measurements constitute evidence that vacuum fluctuations of the microwave field can affect the dynamics of spins, and, therefore, are a step towards the application of concepts in circuit quantum electrodynamics to individual spins in solids.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 24 August; accepted 11 December 2015.**

**Published online 15 February 2016.**

- Purcell, E. M. Spontaneous emission probabilities at radio frequencies. *Phys. Rev.* **69**, 681 (1946).
- Goy, P., Raimond, J. M., Gross, M. & Haroche, S. Observation of cavity-enhanced single-atom spontaneous emission. *Phys. Rev. Lett.* **50**, 1903–1906 (1983).
- Heinzen, D. J., Childs, J. J., Thomas, J. E. & Feld, M. S. Enhanced and inhibited visible spontaneous emission by atoms in a confocal resonator. *Phys. Rev. Lett.* **58**, 1320–1323 (1987).

- Yamamoto, Y., Machida, S., Horikoshi, Y., Igeta, K. & Bjork, G. Enhanced and inhibited spontaneous emission of free excitons in GaAs quantum wells in a microcavity. *Opt. Commun.* **80**, 337–342 (1991).
- Gérard, J. M. et al. Enhanced spontaneous emission by quantum boxes in a monolithic optical microcavity. *Phys. Rev. Lett.* **81**, 1110–1113 (1998).
- Butler, M. C. & Weitekamp, D. P. Polarization of nuclear spins by a cold nanoscale resonator. *Phys. Rev. A* **84**, 063407 (2011).
- Sinova, J. & Žutić, I. New moves of the spintronics tango. *Nature Mater.* **11**, 368–371 (2012).
- Ladd, T. D. et al. Quantum computers. *Nature* **464**, 45–53 (2010).
- Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance* (John Wiley & Sons, 2001).
- Robledo, L. et al. High-fidelity projective read-out of a solid-state spin quantum register. *Nature* **477**, 574–578 (2011).
- Pla, J. J. et al. A single-atom electron spin qubit in silicon. *Nature* **489**, 541–545 (2012).
- Shapiro, M. G. et al. Directed evolution of a magnetic resonance imaging contrast agent for noninvasive imaging of dopamine. *Nature Biotechnol.* **28**, 264–270 (2010).
- Wood, C. J., Borneman, T. W. & Cory, D. G. Cavity cooling of an ensemble spin system. *Phys. Rev. Lett.* **112**, 050501 (2014).
- Sleator, T., Hahn, E. L., Hilbert, C. & Clarke, J. Nuclear-spin noise. *Phys. Rev. Lett.* **55**, 1742–1745 (1985).
- Hahn, E. Spin echoes. *Phys. Rev.* **80**, 580–594 (1950).
- Zhou, X. et al. High-gain weakly nonlinear flux-modulated Josephson parametric amplifier using a SQUID array. *Phys. Rev. B* **89**, 214517 (2014).
- Bienfait, A. et al. Reaching the quantum limit of sensitivity in electron spin resonance. *Nature Nanotechnol.* <http://dx.doi.org/10.1038/nnano.2015.282> (2015).
- Feher, G. Electron spin resonance experiments on donors in silicon. I. Electronic structure of donors by the electron nuclear double resonance technique. *Phys. Rev.* **114**, 1219–1244 (1959).
- Wolfowicz, G. et al. Decoherence mechanisms of <sup>209</sup>Bi donor electron spins in isotopically pure <sup>28</sup>Si. *Phys. Rev. B* **86**, 245301 (2012).
- Morley, G. W. et al. The initialization and manipulation of quantum information stored in silicon by bismuth dopants. *Nature Mater.* **9**, 725–729 (2010).
- George, R. E. et al. Electron spin coherence and electron nuclear double resonance of Bi donors in natural Si. *Phys. Rev. Lett.* **105**, 067601 (2010).
- Schweiger, A. & Jeschke, G. *Principles of Pulse Electron Paramagnetic Resonance* **225** (Oxford Univ. Press, 2001).
- Feher, G. & Gere, E. A. Electron spin resonance experiments on donors in silicon. II. Electron spin relaxation effects. *Phys. Rev.* **114**, 1245–1256 (1959).
- Bloembergen, N. & Pound, R. V. Radiation damping in magnetic resonance experiments. *Phys. Rev.* **95**, 8–12 (1954).
- Wood, C. J. & Cory, D. G. Cavity cooling to the ground state of an ensemble quantum system. Preprint at <http://arXiv.org/abs/1506.03007> (2015).
- Feher, G., Gordon, J. P., Buehler, E., Gere, E. A. & Thurmond, C. D. Spontaneous emission of radiation from an electron spin system. *Phys. Rev.* **109**, 221–222 (1958).
- Temnov, V. V. & Woggon, U. Superradiance and subradiance in an inhomogeneously broadened ensemble of two-level systems coupled to a low-Q cavity. *Phys. Rev. Lett.* **95**, 243602 (2005).
- Carver, T. R. & Slichter, C. P. Polarization of nuclear spins in metals. *Phys. Rev.* **92**, 212–213 (1953).
- Abraham, A. & Goldman, M. Principles of dynamic nuclear polarisation. *Rep. Prog. Phys.* **41**, 395–467 (1978).
- Bloembergen, N., Shapiro, S., Pershan, P. S. & Artman, J. O. Cross-relaxation in spin systems. *Phys. Rev.* **114**, 445–459 (1959).

**Acknowledgements** We acknowledge technical support from P. Sénat, D. Duet, J.-C. Tack, P. Pari and P. Forget, as well as discussions within the Qnantronics group. We acknowledge support of the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013) through grant agreements No. 615767 (CIRQUSS), 279781 (ASCENT) and 630070 (quRAM), and of the C'Nano IdF project QUANTROCRYO. J.J.L.M. is supported by the Royal Society. C.C.L. is supported by the Royal Commission for the Exhibition of 1851. T.S. and C.D.W. were supported by the US Department of Energy under contract DE-AC02-05CH11231.

**Author Contributions** A.B., J.J.P., J.J.L.M. and P.B. designed the experiment. X.Z. and D.V. designed and fabricated the Josephson Parametric Amplifier. C.C.L., C.D.W. and T.S. provided the bismuth-implanted isotopically purified silicon sample. A.B., J.J.P. and Y.K. fabricated the sample and performed the measurements. A.B., J.J.P., Y.K., J.J.L.M. and P.B. analysed the data. J.J.L.M., D.E., D.V. and P.B. supervised the project. A.B., J.J.P., Y.K., M.S., D.V., D.E., J.J.L.M. and P.B. contributed to writing the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.B. ([patrice.bertet@cea.fr](mailto:patrice.bertet@cea.fr)).



## METHODS

**Bismuth donors in silicon.** Bismuth donors in silicon have the following isotropic spin Hamiltonian<sup>31</sup>:  $\hat{H}/h = \mathbf{B} \cdot (\gamma_e \mathbf{S} \otimes \mathbf{I} - \gamma_n \mathbf{I} \otimes \mathbf{I}) + \mathbf{A} \mathbf{S} \cdot \mathbf{I}$ , in which the electronic gyromagnetic ratio  $\gamma_e/(2\pi) = 27.997 \text{ GHz T}^{-1}$ , the nuclear gyromagnetic ratio  $\gamma_n/(2\pi) = 6.9 \text{ MHz T}^{-1}$  and the hyperfine coupling constant  $A/h = 1.475 \text{ GHz}$ . For a weak static field  $\mathbf{B}_0$  ( $B_0 \lesssim 50 \text{ mT}$ ) oriented along  $x$ , the eigenstates of the total angular momentum  $\mathbf{F} = \mathbf{S} + \mathbf{I}$  and its projection  $m_F$  along  $\mathbf{B}_0$  represent good quantum numbers for the 20 electro-nuclear energy states of the Bi:Si system. These eigenstates can be grouped in an  $F = 4$  ground and an  $F = 5$  excited multiplet separated by a frequency of  $(I + 1/2)A/h = 7.35 \text{ GHz}$  in zero-field (see Fig. 1d). Transitions between states that verify  $\Delta F \Delta m_F = \pm 1$  can be excited with a field orientated along  $y$  (or  $z$ ) because their associated matrix element  $\langle F, m_F | S_y | F + 1, m_F \pm 1 \rangle = \langle F, m_F | S_z | F + 1, m_F \pm 1 \rangle$  has the same magnitude as an ideal electronic spin 1/2 transition  $\langle m_s | S_y | m_s \rangle = 0.5$ . Only the ten transitions with a matrix element greater than 0.25 are shown in Fig. 2a. Characteristics for the transitions probed by our resonators are given in Extended Data Table 1.

**Single-spin coupling to the resonator.** The spin–resonator interaction is described by a Jaynes–Cummings Hamiltonian,  $\hbar g(a^\dagger \sigma_- + a \sigma_+)$ , in which  $a$  ( $a^\dagger$ ) is the field annihilation (creation) operator,  $\sigma_-$  ( $\sigma_+$ ) is the spin lowering (raising) operator and  $g$  is the spin–resonator coupling strength. For the Bi:Si transitions,  $|F, m_F\rangle \leftrightarrow |F + 1, m_F - 1\rangle$  probed by the resonators,  $g$  can be expressed as  $g = \gamma_e \langle F, m_F | S_x | F + 1, m_F - 1 \rangle \|\delta \mathbf{B}_\perp\|$  (ref. 17), in which  $\delta \mathbf{B}_\perp$  is the component of the resonator-field vacuum fluctuations orthogonal to  $\mathbf{B}_0$ . Considering the orientations for  $\mathbf{B}_0$  and  $\delta \mathbf{B}$  shown in Fig. 1b, we obtain equation (2)

$$g(\theta) = \gamma_e \langle F, m_F | S_x | F + 1, m_F - 1 \rangle \sqrt{\delta B_{1,y}^2 \cos(\theta)^2 + \delta B_{1,z}^2}$$

To estimate the distribution of the coupling constant  $g(0)$  for a given transition, we need to estimate the vacuum-field fluctuations  $\delta \mathbf{B}_1$  in the spin ensemble region. This is achieved using the COMSOL software and assuming a non-homogeneous current-density distribution in the superconducting aluminium wire<sup>32</sup>. The total current flowing through the wire cross-section is  $\delta i = \omega_0 \sqrt{\hbar/2Z_0}$ , in which  $Z_0 = \sqrt{L/C}$  is the resonator impedance, determined to be  $44 \Omega$  via electromagnetic simulations realized in CST Microwave Studio. In all the work presented in the main text, the measurements were done on the low-field peak of transition  $|F, m_F\rangle = |4, -4\rangle \leftrightarrow |5, -5\rangle$ , which has been attributed to spins residing under the wire. From the spin implantation profile (see Fig. 1b) and the spatial dependence of the microwave field  $\delta \mathbf{B}_1$  restrained to the area under the wire ( $|y| < 2.5 \mu\text{m}$ ), the relevant coupling-constant distribution can be extracted. Doing so yields a very asymmetric coupling distribution that is sharply peaked around  $g/(2\pi) = 56 \text{ Hz}$  with a 2-Hz full-width at half-maximum for the transition  $|4, -4\rangle \leftrightarrow |5, -5\rangle$  with  $\theta = 0$ . A more detailed derivation of the coupling constant and its estimate at  $\theta = 0^\circ$  is available in ref. 17.

**Average intra-cavity photon number  $\bar{n}$ .** The average intra-cavity photon number  $\bar{n}$  of a pulse of power  $P_{\text{in}}$  at the cavity input is evaluated as  $\bar{n} = \frac{4\kappa_1 P_{\text{in}}}{\hbar\omega_0(\kappa_1 + \kappa_2 + \kappa_L)^2}$ ,

in which  $\kappa_1$  and  $\kappa_2$  are the couplings to the input and output antennas and  $\kappa_L$  represents the internal losses of the resonator. From a previous calibration of the experimental setup<sup>17</sup>, we estimate that we can determine  $P_{\text{in}}$  with an accuracy of approximately 1 dB. The values of  $\kappa_1$ ,  $\kappa_2$  and  $\kappa_L$  are determined experimentally by measuring each element of the resonator scattering matrix and fitting to the well-known input–output formulae<sup>33</sup>; see Extended Data Table 2.

**Data acquisition and echo signal.** The full description of the experimental set-up is available in ref. 17. The use of a Josephson parametric amplifier allows us to reach a quantum-limited sensitivity. In addition to the Hahn echo sequence, we use a Carr–Purcell–Meiboom–Gill sequence<sup>34</sup> for every echo acquisition. For all  $A_Q$  data points presented in this work, 10  $\pi$  pulses are added after the first echo to recover 10 extra echoes, which are subsequently averaged to boost the signal-to-noise ratio. This scheme allows us to acquire data in single-shot read-out. Each  $A_Q$  data point is a single-shot measurement and the error bars are determined by the variance of a pool of at least  $n = 200$  measurements, taken in similar conditions.

**Experimental determination of  $T_1$  at resonance.** The inversion–recovery sequence is used to measure the spin relaxation time  $T_1$ ; see Fig. 2d. Spin polarization is measured with the following Hahn echo sequence: 50- $\mu\text{s}$ -long  $\pi/2$  pulse, delay  $\tau = 500 \mu\text{s}$ , and 100- $\mu\text{s}$ -long  $\pi$  pulse. The pulse durations were chosen such that only spins within a narrow spectral range were detected, producing a well-defined Purcell-limited  $T_1$ . Indeed, because the probed ensemble of spins has a larger linewidth  $\Delta\omega = 2 \text{ MHz}$  than do our resonators, the signal emitted during the spin-echo comes from a subset of the ensemble of spins, with a frequency spectrum at least as large as the resonator bandwidth. Spins probed at the edges of the bandwidth of the resonator will have longer Purcell relaxation times; for instance, those detuned by  $\delta = \kappa$  have an expected Purcell relaxation time that

is five times slower than the  $T_1$  time expected at perfect resonance; see Extended Data Fig. 1a. The contribution of those spins with a longer decay time to the signal will result in an averaging effect, meaning that the measured  $T_1$  will be erroneously longer than predicted.

To suppress this effect, we reduce the bandwidth of the read-out sequence to collect signal only from spins very close to the resonance. The response function of a pulse of length  $t_p$  incident on a cavity with bandwidth  $\kappa$  at frequency  $\omega_0$  is expressed as

$$\mathcal{R}(\omega) = [2\text{sinc}(t_p(\omega - \omega_0)/2)]^2 \mathcal{R}_{\text{cav}}(\omega) = \frac{[2\text{sinc}(t_p(\omega - \omega_0)/2)]^2}{1 + 4(\omega - \omega_0)^2/\kappa^2}$$

in which  $\text{sinc}(x) = \sin(x)/x$ . As shown in Extended Data Fig. 1a, for the narrowest bandwidth  $\kappa/(2\pi) = 23 \text{ kHz}$  of resonator A, pulses of  $5 \mu\text{s}$  are heavily filtered by the resonator and have the same bandwidth, whereas 100- $\mu\text{s}$ -long pulses have a reduced bandwidth of approximately 10 kHz. In case of 100- $\mu\text{s}$ -long excitation pulses, the Rabi frequency is such that only spins with  $|\delta|/(2\pi) \leq 5 \text{ kHz}$  will contribute to the signal. This corresponds to a dispersion of only 5% for the expected Purcell relaxation times, which is negligible. To illustrate the averaging effect, two inversion–recovery curves are shown in Extended Data Fig. 1b, with readout pulses of  $5 \mu\text{s}$  and  $100 \mu\text{s}$ . The former yields  $T_1 = 0.65 \text{ s}$ , which is a factor two higher than predicted by the Purcell effect, whereas the latter yields the expected value  $T_1 = 0.35 \text{ s}$ .

Therefore, Fig. 2d shows an inversion–recovery sequence that has a read-out echo sequence with a narrow bandwidth ( $t_\pi = 100 \mu\text{s}$ ,  $t_{\pi/2} = t_\pi/2$ ) to suppress contributions from spins with a lower decay rate, and an inversion pulse with a large bandwidth ( $t_\pi = 5 \mu\text{s}$ ) to maximize the efficiency of the inversion.

Given that the spin energy relaxation time  $T_1$  is of the order of 1 s, we choose a repetition rate  $\gamma_{\text{rep}}$  that is sufficiently low to allow full relaxation of the spins between successive inversion–recovery sequences:  $\gamma_{\text{rep}} = 0.04 \text{ Hz}$ .

**Experimental determination of spin–cavity–detuning-dependent relaxation rates.** The spins are detuned from the cavity by applying an additional bias pulse on one of the Helmholtz coils used to apply the static field  $\mathbf{B}_0$ . The extra bias pulse is output by a pulse generator with  $50 \Omega$  output impedance placed in parallel to the d.c. supply of one of the Helmholtz coils. To minimize the effect of transients due to the 1-Hz bandwidth of the coils, buffer times of 1 s are added after ramping the coil up and down. To limit the loss of signal during these buffer times, we use an angle  $\theta = 45^\circ$  and work with resonator B in order to have a longer  $T_1(0) = 1.68 \text{ s}$ . Applying a magnetic-field pulse to a single coil instead of both coils perturbs  $\theta$  by at most  $4^\circ$ . The value of  $T_1(0)$  was measured with inversion recovery. All the data presented in Fig. 3 and in Extended Data Fig. 2 were acquired in a separate run. The quality factor of resonator B decreased from  $Q = 1.07 \times 10^5$  to  $Q = 8.9 \times 10^4$  owing to slightly higher losses, yielding the resonator bandwidth  $\kappa/(2\pi) = 82 \text{ kHz}$ .

To observe the long relaxation times, such as those measured in Fig. 3, inversion recovery is not an ideal method. When the spin linewidth is broader (about 20 times) than the excitation bandwidth and when the thermalization time is very long, one can observe polarization mixing mechanisms<sup>35,36</sup>, spectral and spatial spin diffusion being the most relevant to our case, because the spin system is composed of only one species. If we try to measure the relaxation from spins that have been detuned by an amount  $\delta/(2\pi) = (\omega_s - \omega_0)/(2\pi) = 3.8 \text{ MHz}$  during a lapse of time  $T$  with an inversion–recovery sequence (Extended Data Fig. 2a), then we observe a double-exponential relaxation (Extended Data Fig. 2d, green), which we attribute to the existence of a spin-diffusion mechanism.

Spin diffusion is prevented by suppressing any polarization gradient along the spin line, which leads us to use a saturation–recovery scheme instead of inversion recovery. The simplest saturation–recovery scheme (Extended Data Fig. 2b) consists of sending a strong microwave tone that results in the saturation of the line, producing an incoherent mixed state with the population evenly distributed between excited and ground states. Nevertheless, a relaxation time measured using this scheme still yields a double-exponential decay (Extended Data Fig. 2d, orange), with time constants similar to those for the inversion recovery case. This implies that the saturation of the line is insufficient.

To improve the saturation, we can sweep the magnetic field during the saturation pulse to bring different subsets of the spin line to resonance and realize a full saturation. The adopted sweep scheme is shown in Extended Data Fig. 2c. The corresponding relaxation curve fits well to a simple exponential decay (Extended Data Fig. 2d, blue), indicating the suppression of the spin-diffusion effect.

We further check the quality of the saturation by measuring the polarization across the full spin linewidth immediately after saturation. To realize such scans (Extended Data Fig. 2e), we apply the relevant saturation pulse at  $\omega_0$ , then apply a magnetic field pulse  $B_\delta = (\omega_s - \omega_0)/\gamma_e$  and measure the echo signal  $A_Q(\omega_s)$  with a Hahn echo sequence. When no saturation pulse is applied, the measured echo

signal  $A_{Q0}(\omega_s)$  is a measure of the full polarization  $-\langle S_z(\omega_s) \rangle = +1$  (Extended Data Fig. 2e, black curve) and shows the natural spin linewidth. When studying an excitation pulse, the polarization of the spins is  $-\langle S_z(\omega_s) \rangle = A_Q(\omega_s)/A_{Q0}(\omega_s)$ , in which  $A_Q(\omega_s)$  is the measured echo signal. Therefore,  $-\langle S_z(\omega_s) \rangle = -1$  indicates full inversion,  $\langle S_z(\omega_s) \rangle = 0$  indicates saturation and  $-\langle S_z(\omega_s) \rangle = +1$  indicates return to thermal equilibrium. The green, orange and blue curves are taken after a  $\pi$  pulse, after saturation without field sweep and after saturation with field sweep, respectively. At resonance, we expect a change of  $S_z$  from  $-1$  to  $+1$  for a  $\pi$  pulse and from  $-1$  to  $0$  for a saturation pulse. Owing to the coil transient time, all three curves show a partial relaxation. If the saturation was optimal and no partial relaxation was occurring, then we should observe  $S_z = 0$  for any detuning  $\delta$ . For the two saturations (with and without field sweep) studied here, only that with field sweep equally saturates the line. The basic saturation with field sweep has a bandwidth of approximately 250 kHz and the bandwidth of the  $\pi$  pulse is similar to that of the cavity  $\kappa/(2\pi) = 82$  kHz. This finding confirms that spin diffusion is fully suppressed only in a scheme of saturation with field sweep, to yield a simple exponential-decay relaxation. This scheme is used to measure the 22 relaxation rates at different detunings  $\delta$  shown in Fig. 4.

The global fit shown in Fig. 4c is obtained using  $[T_1(\delta)]^{-1} = \Gamma_p + \Gamma_{NR}$ , which may be expressed as  $[T_1(0)]^{-1} [1 + 4(\delta/\kappa)^2]^{-1} + \Gamma_{NR}$  to involve only experimentally determined parameters. Indeed,  $\kappa$  is precisely determined by measuring the quality factor of the resonator at low power, and  $T_1(0)$  is determined by an inversion-recovery sequence, as mentioned above. The parameter  $\delta$  was determined via precise calibration of the coil pulse. Therefore, the only remaining free parameter in the fit is  $\Gamma_{NR}$ , yielding  $\Gamma_{NR}^{-1} = 1,600$  s. The error bars come from the accuracy of the fits of the relaxation rates.

**Practical considerations for the application of cavity-induced relaxation in magnetic resonance.** The experiments described in the main text take place at low magnetic field ( $B_0 < 10$  mT) and low temperature ( $T \approx 20$  mK) using a dilution refrigerator. These are unusual conditions for magnetic resonance measurements; however, as we discuss below, with straightforward modifications, cavity-induced relaxation could be observed in other environments, broadening the class of spin systems that could be used.

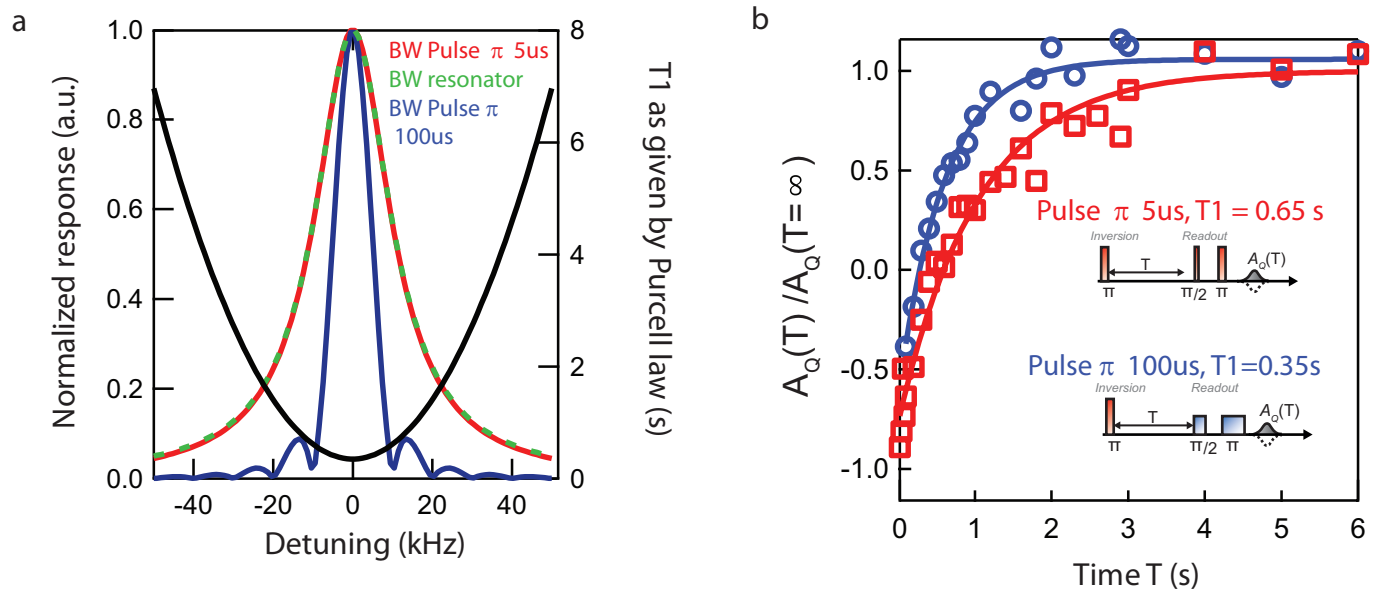
Superconducting microresonators can withstand large magnetic fields (up to approximately 1 T) while maintaining a large quality factor ( $Q_i \approx 2 \times 10^5$ )<sup>37–39</sup> if they are patterned in metals such as Nb, NbN or NbTiN, instead of Al, as we have used. The use of these alternative metals would enable our results to be applied to a much larger class of spin systems, including typical electron spins with  $g \approx 2$ . Similar observations can be made for temperature: Nb, NbN and NbTiN have a higher critical temperature than does Al, which would permit the use of temperatures of 1–4 K (accessible with conventional liquid helium cryostats). However, temperature is important for reasons other than helping to maintain a small  $\kappa$ , because the Purcell effect brings spins into thermal equilibrium with the cavity field; for example, at the microwave frequencies used in our experiments (7.3 GHz),

temperatures below 70 mK are required for a spin polarization of  $>99\%$ . Higher temperatures could be used at the cost of the degree of spin polarization, but this issue could be addressed by moving to higher frequencies. A third factor when considering the operating temperature is that cavity-induced relaxation can only be exploited when it dominates over intrinsic processes such as spin-lattice relaxation. For most spin systems, this requirement translates into temperatures similar to those of liquid-helium.

The possibility of cavity-induced relaxation with conventional electron spin systems might lead to applications other than those that benefit from a faster return to thermal equilibrium to increase signal averaging rates. In particular, we consider the possibility of cavity-assisted dynamic nuclear polarization, via either the so-called solid effect or the Overhauser effect, which was recently observed in solids<sup>40</sup>. With the solid effect, the equilibrium polarization of a nuclear spin of frequency  $\omega_n$  coupled to an electron spin of frequency  $\omega_e$  is enhanced by irradiating it with microwaves at  $\omega_e + \omega_n$ , provided the electron spins return quickly enough to equilibrium. Tuning a cavity on resonance with the electron spin transition at  $\omega_e$  could provide an alternative relaxation mechanism to phonons, thereby avoiding, for example, phonon bottleneck effects and/or mitigating the need to apply large magnetic fields. With the Overhauser effect, saturating the spin transition by applying microwaves at  $\omega_s$  enhances the nuclear spin polarization because of the existence of electron-nuclear spin cross-relaxation processes, which could be enhanced by tuning a cavity at  $\omega_e - \omega_n$ .

**Sample size.** No statistical methods were used to predetermine sample size.

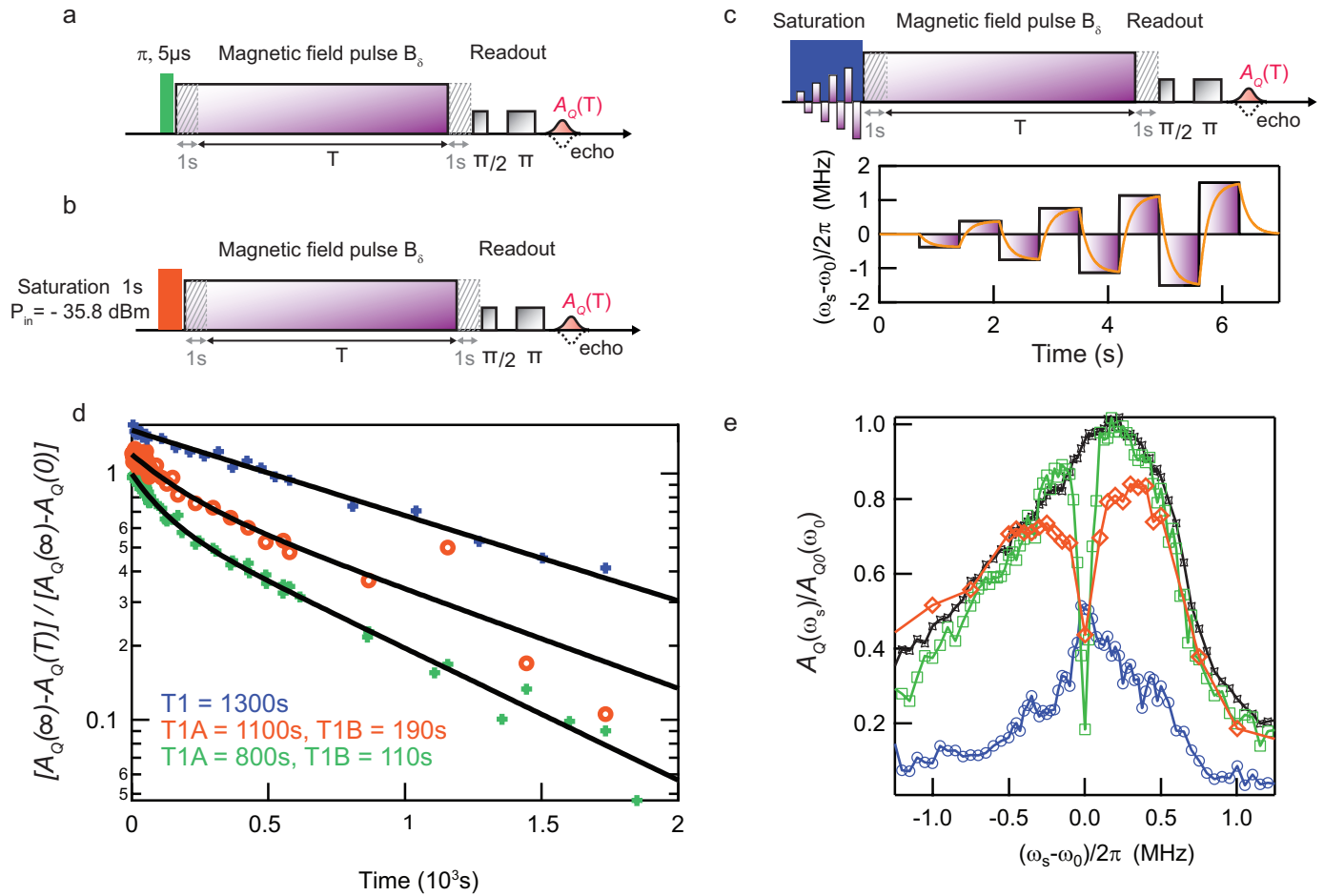
31. Wolfowicz, G. *et al.* Atomic clock transitions in silicon-based spin qubits. *Nature Nanotechnol.* **8**, 561–564 (2013).
32. Van Duzer, T. & Turner, C. W. *Principles of Superconductive Devices and Circuits* 2nd edn (Prentice-Hall PTR, 1999).
33. Palacios-Laloy, A. *Superconducting Qubit in a Resonator: Test of the Leggett-Garg Inequality and Single-shot Readout*. PhD thesis, Université Pierre et Marie Curie — Paris VI (2010).
34. Mentink-Vigier, F. *et al.* Increasing sensitivity of pulse EPR experiments using echo train detection schemes. *J. Magn. Reson.* **236**, 117–125 (2013).
35. Bloembergen, N. On the interaction of nuclear spins in a crystalline lattice. *Physica* **15**, 386–426 (1949).
36. Abragam, A. *Principles of Nuclear Magnetism* Ch. IX (Oxford Univ. Press, 1983).
37. de Graaf, S. E., Davidovikj, D., Adamyan, A., Kubatkin, S. E. & Danilov, A. V. Galvanically split superconducting microwave resonators for introducing internal voltage bias. *Appl. Phys. Lett.* **104**, 052601 (2014).
38. Wisby, I. *et al.* Coupling of a locally implanted rare-earth ion ensemble to a superconducting micro-resonator. *Appl. Phys. Lett.* **105**, 102601 (2014).
39. Samkharadze, N. *et al.* High kinetic inductance superconducting nanowire resonators for circuit QED in a magnetic field. Preprint at <http://arXiv.org/abs/1511.01760> (2015).
40. Can, T. V. *et al.* Overhauser effects in insulating solids. *J. Chem. Phys.* **141**, 064202 (2014).



**Extended Data Figure 1 | Effect of excitation-pulse bandwidth on the measurement of  $T_1$ .** **a**, The red and blue lines shown the computed pulse bandwidth ('normalized response') for a 5- $\mu$ s  $\pi$  pulse and a 100- $\mu$ s  $\pi$  pulse, respectively, incident on a cavity with  $\kappa/(2\pi) = 23$  kHz (green dashes). To illustrate the averaging effect of the pulse bandwidth on  $T_1$  measurements, the expected Purcell  $T_1$  curve (black line) as a function of spin-cavity detuning is plotted on the right axis, with  $T_1(0) = 0.35$  s

and  $\kappa/(2\pi) = 23$  kHz. **b**,  $T_1$  measurements for two different  $\pi$ -pulse lengths (see insets), measured on resonance with resonator A. Spin polarization is measured with a Hahn echo sequence and  $A_Q$  is rescaled by its value for  $T \gg T_1$  (' $A_Q(T = \infty)$ '). Symbols are data and solid lines are exponential fits. The 100- $\mu$ s  $\pi$  pulse (blue) yields  $T_1 = 0.35$  s, which is in agreement with the Purcell rate. The 5- $\mu$ s  $\pi$  pulse (red) yields  $T_1 = 0.65$  s, a factor of two greater than the accurate value.





**Extended Data Figure 2 | Spectral spin diffusion.** **a–c**,  $T_1$  measurement sequence when spins are detuned from the cavity by applying a magnetic field  $B_\delta$ , providing a detuning of  $\delta = \omega_s - \omega_0 = 2\pi\gamma_{\text{eff}}B_\delta$ , with  $\gamma_{\text{eff}} = df/dB(B_0)$  the effective gyromagnetic ratio, evaluated as the derivative of  $f = 2\pi\omega_s$  with respect to the applied magnetic field  $B$  at a given magnetic field  $B_0$ . In **a**, a 5- $\mu$ s  $\pi$  pulse is used to realize an inversion-recovery sequence; in **b**, a 1-s-long strong microwave pulse sent at cavity resonance is used to realize a saturation-recovery sequence; in **c**, a magnetic field scan (bottom panel) is used in addition to a 6-s-long strong microwave pulse to realize a saturation-recovery sequence. The expected magnetic field profile due to the coil filtering, assuming that the coil is an order-one low-pass filter with a bandwidth of 1 Hz, is shown in orange (**c**, bottom panel). **d**,  $T_1$  measurements for sequences shown in **a** (green),

**b** (red) and **c** (blue) for  $\delta/(2\pi) = 3.8$  MHz. The fits (black lines) to the green and red data have a double-exponential decay, whereas the fit to the blue data is a simple exponential. We attribute the double-exponential decay (with extracted characteristic times  $T_{1A}$  and  $T_{1B}$ ) to spin diffusion. **e**, Spectral profiles of the excitation pulse sequences shown in **a** (green), **b** (red) and **c** (blue). The sequence is as follows: send the excitation pulse, detune the spins and measure  $A_Q(\omega_s)$ . The black line is the reference profile without any excitation pulse, yielding the reference polarization  $\langle S_z(\omega_s) \rangle = -A_{Q0}(\omega_s)/A_{Q0}(\omega_0) = -1$ . When an excitation pulse is sent, we can access  $\langle S_z(\omega_s) \rangle = -A_Q(\omega_s)/A_{Q0}(\omega_0)$ . To conserve the line shape profile, we plotted  $A_Q(\omega_s)/A_{Q0}(\omega_0)$  instead of  $A_Q(\omega_s)/A_{Q0}(\omega_s)$ . Neither the  $\pi$  profile nor the saturation profiles reach the full inversion +1 or the full saturation 0 at resonance; this is an artefact due to the coil transient time.

Extended Data Table 1 | Relevant Bi:Si transitions and their characteristics

Transition	$df/dB$	$\langle F, m_F   S_x   F + 1, m_F - 1 \rangle$
$ 4, -4\rangle \leftrightarrow  5, -5\rangle$	$-25.1 \text{ GHz/T}$	0.47
$ 4, -3\rangle \leftrightarrow  5, -4\rangle$	$-19.2 \text{ GHz/T}$	0.42
$ 4, -2\rangle \leftrightarrow  5, -3\rangle$	$-13.5 \text{ GHz/T}$	0.37

Extended Data Table 2 | Resonator characteristics

Resonator	$\omega_r/2\pi$	Q	$\kappa_1$ (s <sup>-1</sup> )	$\kappa_2$ (s <sup>-1</sup> )	$\kappa_L$ (s <sup>-1</sup> )
A	7.2467 GHz	$3.2 \times 10^5$	$1.3 \times 10^4$	$5.8 \times 10^4$	$7.5 \times 10^4$
B	7.3054 GHz	$1.1 \times 10^5$	$3.6 \times 10^4$	$3.1 \times 10^5$	$8.2 \times 10^4$



# Condensation on slippery asymmetric bumps

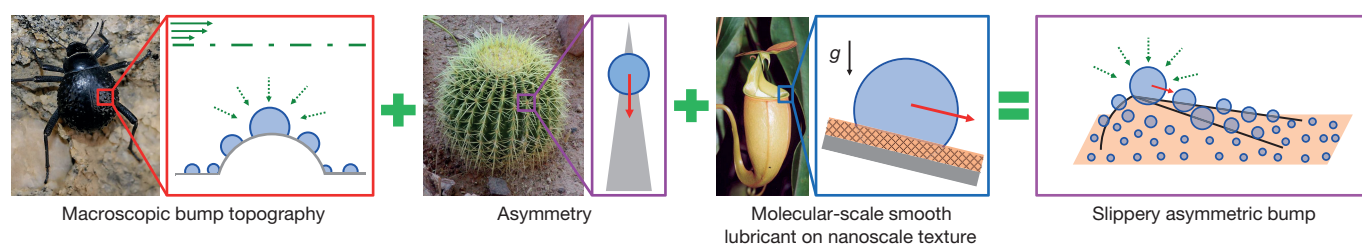
Kyoo-Chul Park<sup>1,2</sup>, Philseok Kim<sup>2</sup>, Alison Grinthal<sup>1</sup>, Neil He<sup>1</sup>, David Fox<sup>1</sup>, James C. Weaver<sup>2</sup> & Joanna Aizenberg<sup>1,2,3</sup>

Controlling dropwise condensation is fundamental to water-harvesting systems<sup>1–3</sup>, desalination<sup>4</sup>, thermal power generation<sup>4–8</sup>, air conditioning<sup>9</sup>, distillation towers<sup>10</sup>, and numerous other applications<sup>4,5,11</sup>. For any of these, it is essential to design surfaces that enable droplets to grow rapidly and to be shed as quickly as possible<sup>4–7</sup>. However, approaches<sup>4–8,10–21</sup> based on microscale, nanoscale or molecular-scale textures suffer from intrinsic trade-offs that make it difficult to optimize both growth and transport at once. Here we present a conceptually different design approach—based on principles derived from Namib desert beetles<sup>3,22–24</sup>, cacti<sup>25</sup>, and pitcher plants<sup>17,26</sup>—that synergistically combines these aspects of condensation and substantially outperforms other synthetic surfaces. Inspired by an unconventional interpretation of the role of the beetle's bumpy surface geometry in promoting condensation, and using theoretical modelling, we show how to maximize vapour diffusion flux<sup>20,27,28</sup> at the apex of convex millimetric bumps by optimizing the radius of curvature and cross-sectional shape. Integrating this apex geometry with a widening slope, analogous to cactus spines, directly couples facilitated droplet growth with fast directional transport, by creating a free-energy profile that drives the droplet down the slope before its growth rate can decrease. This coupling is further enhanced by a slippery, pitcher-plant-inspired nanocoating that facilitates feedback between coalescence-driven growth and capillary-driven motion on the way down. Bumps that are rationally designed to integrate these mechanisms are able to grow and transport large droplets even against gravity and overcome the effect of an unfavourable temperature gradient. We further observe an unprecedented sixfold-higher exponent of growth rate, faster onset, higher steady-state turnover rate, and a greater volume of water collected compared to other surfaces. We envision that this fundamental understanding and rational design strategy can be applied to a wide range of water-harvesting and phase-change heat-transfer applications.

The central concepts for integrating the growth and transport of water droplets are derived from a combination of strategies used by

three distinct biological examples—Namib desert beetles<sup>3,22–24</sup>, cacti<sup>25</sup>, and *Nepenthes* pitcher plants<sup>26</sup>—as summarized in Fig. 1. Several desert beetle species harvest water using their bumpy backs, but the topography of these surfaces has been overlooked, primarily because most previous studies attributed preferential condensation to surface chemistry (hydrophilic bumps with hydrophobic surroundings) and discounted convex topography as inferior to concave, on the basis of research into microscale and nanoscale textures<sup>13,14,22,29,30</sup>. However, the beetle bumps are large (millimetres across), and recent studies have reported that the entire bumpy surfaces are homogeneously covered with hydrophobic wax<sup>23,24</sup>, thus questioning the role of localized surface chemistry in promoting condensation. We considered instead that even in the absence of localized chemical patterning or microscale/nanoscale textures, the specific geometry of convex millimetre-sized surface structures alone could facilitate condensation, and therefore the topography of synthetic bumpy surfaces could potentially be designed to optimize fast, localized droplet growth by focusing vapour diffusion flux<sup>20,27,28</sup> at the apex. We further hypothesized that fast growth of millimetre-sized droplets could be coupled to rapid directional turn-over by integrating optimized apex geometry with an asymmetric slope analogous to that used by cactus spines to guide capillary-driven transport of harvested water drops. Lastly, the negligible friction of the slippery coating of pitcher plants inspired us to coat the bumps with molecularly smooth lubricant immobilized on nanotexture to facilitate these topography-based mechanisms.

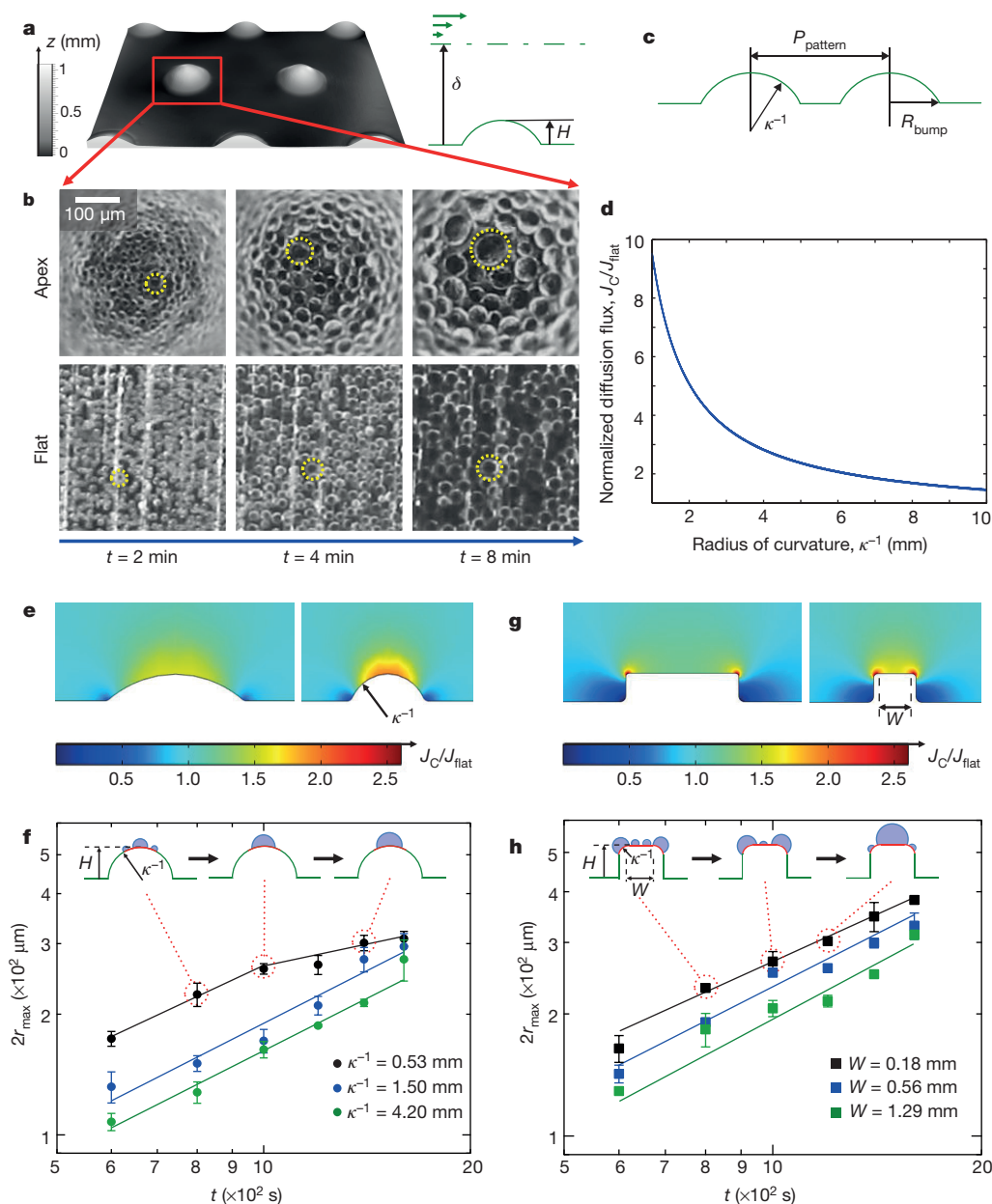
To test the idea that droplets grow faster on convex millimetre-sized surface structures by focused diffusion flux, we measured the diameter of the largest droplet growing on the apex of a simple spherical-cap-shaped bump shown in Fig. 2a (see Supplementary Methods for the fabrication process, Supplementary Tables 1 and 2 for surface characterization, and Supplementary Fig. 1 for experimental setup). In this test, the depletion layer thickness  $\delta$  (of the order of 10 nm), where convection normal to the surface is negligible, is much greater than the height of the cap ( $H \approx 0.8 \pm 0.25$  mm), making diffusion of water vapour the dominant mass transport mechanism<sup>28</sup>. As shown in Fig. 2b, the diameter of



**Figure 1 | An overview of our approach.** The convex millimetric bump topography of water-harvesting desert beetles inspired the rational design of bumps that optimize fast, localized droplet growth by focusing vapour diffusion flux (green dotted arrows) at the apex. We further designed the bump to include an asymmetric slope analogous to that of cactus spines for capillary-guided directional transport of harvested droplets, and a pitcher-plant-inspired molecularly smooth lubricant immobilized on nanotexture. The green dash-dotted line represents the depletion layer. The green, black

and red solid arrows represent the directions of convection, gravity  $g$  and drop transport, respectively. Approximate species dimensions are: desert beetle, 1.5 cm long; cactus, 20 cm in diameter; pitcher plant, 15 cm tall. Images are adapted with permission from Wikimedia Commons (desert beetle photograph by Hans Hillewaert, cactus photograph by Stan Shebs), and ref. 26 (pitcher plant photograph, copyright (2004) National Academy of Sciences, USA).

<sup>1</sup>John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. <sup>2</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts, USA. <sup>3</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA.



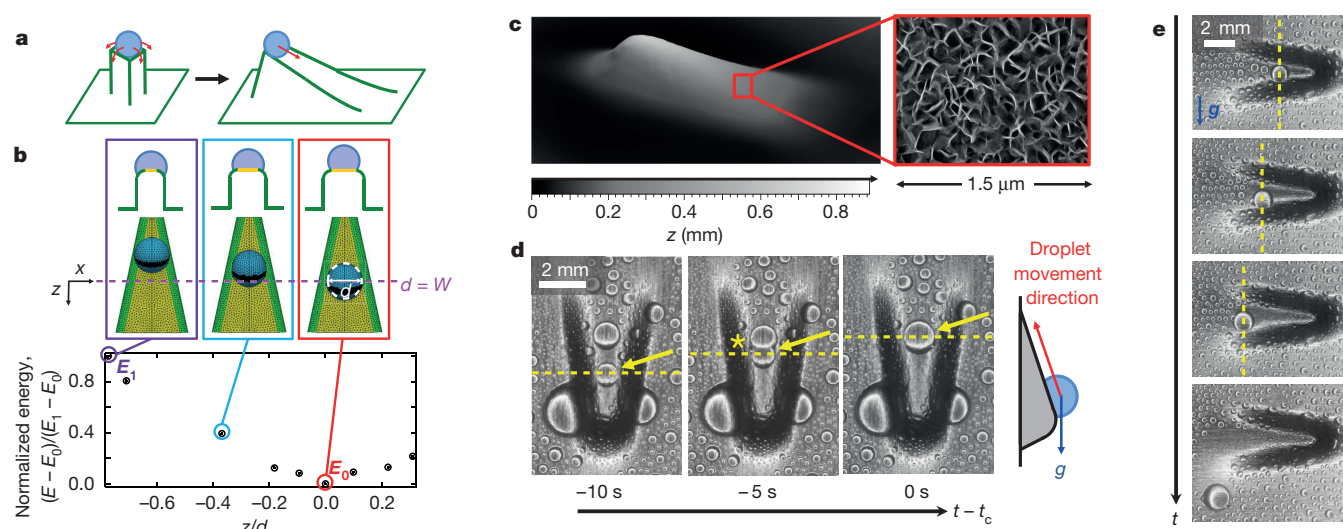
**Figure 2 | Design of, and droplet growth on, convex bumps.**

**a**, A profilometer image of a hexagonal array of millimetre-sized spherical-cap-shaped bumps (left), and the location of the plane (green dash-dotted line, right), below which diffusion is the dominant mechanism of mass transport (depletion layer thickness  $\delta \gg H$ ). **b**, Time-lapsed images of droplets growing on the apex of the bumps (top row) compared to a flat region with the same height  $H$  (bottom row). Yellow circles indicate the largest droplet in each image (see Supplementary Video 1). **c**, A schematic illustration indicating the radius of curvature  $\kappa^{-1}$ , periodicity  $P_{\text{pattern}}$ , and half-bump width  $R_{\text{bump}}$ . **d**, Predicted diffusion flux  $J_C$  on the apex

of a spherical bump as a function of the radius of curvature  $\kappa^{-1}$  (see Supplementary Fig. 5 for the analytical model used in the derivation of  $J_C$ ). ( $J_{\text{flat}}$  is the predicted diffusion flux on a flat surface with the same depletion layer thickness.) **e**, Numerically calculated intensity profile of diffusion flux (COMSOL-Multiphysics). **f**, Time-dependent droplet growth on bumps with decreasing radii of curvature.  $2r_{\text{max}}$  is the averaged diameter of the three largest droplets. **g**, Numerical calculation of diffusion flux and quantitative analysis of droplet growth on rectangular bumps of width  $W$ . **h**, Time-dependent droplet growth on bumps with decreasing width. All error bars are 1 s.d.

the largest droplet on the apex of the bump, denoted by yellow dotted circles in the top images, is greater than that of a droplet on the nearby flat regions of the same surface (yellow dotted circles in the bottom images) at each time point (see Supplementary Video 1). Detailed studies were performed to rule out the role of surface roughness, chemistry and temperature differences in promoting such site-specific droplet growth (see the ‘Control Experiments’ section of the Supplementary Information, Supplementary Video 2, and Supplementary Figs 2–4 for discussion). These initial results indicate that the convex macroscopic surface topography—positive radius of curvature  $\kappa^{-1}$  in Fig. 2c—plays an important part in controlling diffusion flux.

To optimize the focused diffusion flux, we developed predictive models that quantify the magnitude and spatial profile of vapour flux as a function of the radius of curvature. A plot of the simple scaling of diffusion flux near the apex of a spherical cap ( $J_C \propto 1/\kappa^{-1}$ ; see the ‘Theoretical modelling’ section of the Supplementary Information and Supplementary Fig. 5 for a detailed derivation) shows that the smaller the radius of curvature, the greater the localized diffusion flux (Fig. 2d). Numerical calculation (Fig. 2e) using the COMSOL-Multiphysics program (<https://www.comsol.com/comsol-multiphysics>) enables us to visualize more precisely the spatial distribution and intensity of diffusion flux on the bump and surrounding flat regions.



**Figure 3 | Capillary-driven transport.** **a**, Transport (red arrows) on a rectangular bump (left) and an asymmetric slope (right). **b**, Energy profile of the asymmetric bump–droplet–vapour system, obtained by finite-element-based numerical calculation (Surface Evolver).  $E_1$  and  $E_0$  represent the calculated energy values of the system in the purple and red boxes, respectively. **c**, A profilometer image (oblique view) showing nanostructure (inset) infused with lubricant, and the tangentially connected bottom slope. **d**, **e**, Time-lapsed optical images of condensed water droplets on an asymmetric bump rotated 180° (**d**) and 90° (**e**) relative to gravity, corresponding to Supplementary Video 3 (starting at

running time  $\sim 19$  s) and Supplementary Video 4 (starting at running time  $\sim 8$  s). The droplet shown by the yellow arrow partially covers the curved border (indicated by a higher reflection that can be seen as a thin bright region between the black sides and the grey flat top) of the asymmetric bump ( $t - t_c = -10$  s, where  $t_c$  is the time of completed coalescence). The yellow asterisk indicates coalescence with another drop. The dotted yellow line tracks the vertical (**d**) or horizontal (**e**) progress of the droplet. See Supplementary Videos 3–5 for more information about the role of a gradually widening slope.

As indicated by the red to yellow colour gradient, the area where the diffusion flux is greater than on the flat region with the same height becomes increasingly concentrated at and around the apex, and its maximum intensity grows stronger, as the radius of curvature decreases.

Consistent with the analytical and numerical models, we experimentally observe the largest droplet diameters at the apex of spherical-cap-shaped surface features that have the smallest radii of curvature (Fig. 2f, Supplementary Table 3). However, the rate of droplet growth on the bump with the smallest radius of curvature begins to slow down at later time points (see black curve in Fig. 2f). This change of slope suggests that the effect of the focused diffusion flux at the apex diminishes when this region (represented by the red colour in the illustration in Fig. 2f) becomes covered by the growing droplet. To maintain the advantages of the small radius of curvature but avoid this decrease in growth rate, we changed the geometry of the apex from spherical to rectangular, with a flat region on top bordered by rounded edges as shown in Fig. 2g and h. This shape allows us to incorporate an even smaller radius of curvature around the perimeter, combined with an additional area of focused flux on the top flat area. For a smaller width, the superposition of diffusion flux focused on these features collectively results in a larger contiguous area of high diffusion flux. Droplets on the rectangular structure, therefore, continue growing for a longer time at a constant growth rate, as shown in Fig. 2h, because the coalescence and movement of the growing droplets to the flat top area of the bump continues to provide fresh sites for re-nucleation and growth.

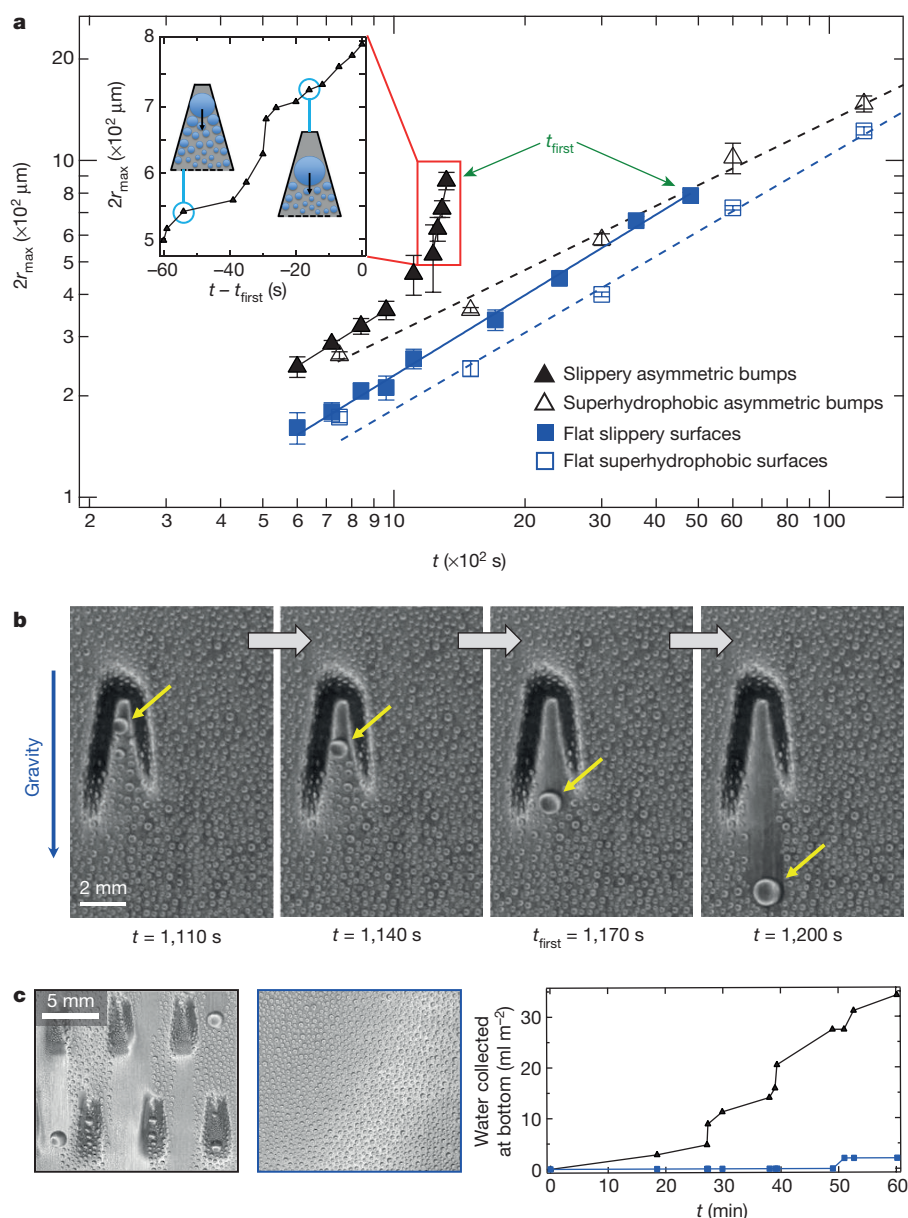
As the growing droplet begins to cover the curved edges, the shape of the rectangular structure—flat with curved borders—also lends itself to a mechanism to transport the droplet directionally, when topographical asymmetry is integrated into the design. Whereas a droplet growing on a rectangular column will eventually fall off in a random direction (Fig. 3a), the addition of a gradually widening slope descending from one side is expected to promote downward motion by enabling the drop to transition to a completely flat surface. As shown by the numerical calculation (Supplementary Fig. 6) the total free energy of the droplet–bump–vapour system is lowest when the droplet is on a completely flat region of such an asymmetric convex structure where the diameter of a droplet  $d$  equals the width of the underlying asymmetric convex

structure  $W$  (Fig. 3b). Simulation (using the Surface Evolver program; <http://facstaff.susqu.edu/brakke/evolver/evolver.html>) suggests that the capillary force resulting from this energy profile on a surface with negligible friction would lead the drop to move down the slope towards the wider flat area (represented by the yellow surface in Fig. 3b) such that it will no longer overlap with the curved regions.

To experimentally validate the latter hypothesis, we fabricated asymmetric bumps with a tangential connection between the descending slope and the surrounding flat regions (Fig. 3c). The negligible friction and pinning assumed in the numerical calculation were achieved by incorporating the pitcher-plant-inspired slippery nanocoating<sup>11,17,18</sup> on top of the entire surface. On the fabricated slippery asymmetric structures, droplets move even against gravity (Fig. 3d and Supplementary Video 3) because in such a system the capillary effect is dominant compared to gravitational effect (as captured by the Bond number—the ratio between gravitational force and capillary force—which is  $(\rho_{\text{water}} - \rho_{\text{air}})gr^2/\gamma_{\text{LV}} \approx 1/7$  at the length scale of the droplet, where  $\rho_{\text{water}}$  and  $\rho_{\text{air}}$  are the density of water and air,  $g$  is the gravitational constant,  $r$  is the radius of the droplet, and  $\gamma_{\text{LV}}$  is the surface tension of the water–vapour interface). In addition, even though the lowest energy point for the moving droplet with a constant diameter is not the bottom of the bump (that is, its widest region) and the droplet might be expected to be pinned upon reaching the point where  $d = W$ , the droplet (indicated by a yellow arrow in Fig. 3d) keeps moving and accelerating as its size grows by coalescing with other small droplets on its path (for example, as indicated by an asterisk in Fig. 3d), thus continuing to satisfy the condition  $d > W$ . This mechanism guides droplet motion solely along the direction determined by the widening slope, regardless of the orientation of the bump relative to gravity, as shown in Fig. 3e and Supplementary Videos 4 and 5.

When surface structures are designed to optimize and integrate all of these mechanisms cooperatively, droplets rapidly grow and start to shed much earlier than on other state-of-the-art surfaces. Figure 4a provides quantitative analysis of the droplet growth as a function of time, comparing the performance of slippery and superhydrophobic surfaces with and without asymmetric bumps. In agreement with earlier studies<sup>6,7,11,17</sup>, flat slippery surfaces outperform superhydrophobic





**Figure 4 | Coupled fast growth and transport.** **a**, Quantitative analysis of growth dynamics and shedding compared to state-of-the-art surfaces. Both slippery asymmetric bumps and superhydrophobic asymmetric bumps show faster localized droplet growth in the early stage ( $t < 10^3 \text{ s}$ ) compared to flat slippery and superhydrophobic control surfaces. Each data point represents the averaged value of the largest droplet's growth on each of three different bumps, or locations in the case of flat surfaces. The red box indicates further enhanced growth rate on slippery asymmetric bumps.  $t_{\text{first}}$  is the average time at which the first three drops

are transported solely by gravity. Each data point in the red box represents the average size of these three drops. The inset shows a magnified view of the fast growth. All error bars are 1 s.d. **b**, Time-lapsed images of droplets condensed on slippery surfaces. **c**, An exemplary array of the slippery asymmetric bumps (left image, black line in the plot) shows a substantially greater volume of water collected at the bottom of the surface (see Supplementary Video 6), compared to the flat slippery surfaces (right image, blue line in the plot).

surfaces in droplet size measured at a given time. Moreover, owing to the higher diffusion flux at the apex of the convex features, both slippery asymmetric bumps (solid black line) and superhydrophobic asymmetric bumps (dotted black line) show faster localized droplet growth in the early stage ( $t < 10^3 \text{ s}$ ) than do their flat controls. Even more interestingly, surfaces with slippery asymmetric bumps show a unique discontinuous behaviour, with a slope of  $\sim 0.82$  at the early stage and  $\sim 6.4$  at the later stage of droplet growth, which is more than six-fold higher than the maximum slope ( $\sim 1$ ) observed in typical droplet growth dynamics<sup>27,28</sup>. Supplementary Figs 7 and 8 demonstrate the importance of combining all the described features of the multi-scale surface structures in achieving successful droplet growth and transport: the absence of any of the design components noticeably slows down

droplet growth dynamics. The accelerated growth (slope of 6.4) captured by the magnified view in the inset of Fig. 4a (see Supplementary Fig. 9 for more information about the accelerated growth of individual droplets compared to other controls) can be interpreted as the feedback between coalescence-driven growth and capillary-driven transport discussed above. As a result, the fast-growing droplets on the slippery asymmetric bumps, which are aligned with gravitational force, are delivered to the bottom of the slope at a size where they can then be transported by gravity, while droplets on the adjacent flat slippery surfaces are still far below the critical shedding diameter (Fig. 4b).

In our experiments, droplets shed from the slippery bumps within  $t_{\text{first}} \approx 10^3 \text{ s}$  (where  $t_{\text{first}}$  is the average time at which the first three drops are transported solely by gravity), whereas droplets on

other state-of-the-art surfaces grow slowly, and shed much later (for example,  $t_{\text{first}} \approx 4 \times 10^3$  s on flat slippery surfaces) or do not shed for more than  $t \approx 10^4$  s (for example, on superhydrophobic surfaces and all other controls shown in Supplementary Figs 7 and 8). Owing to this faster droplet growth and transport performance, which yields a continuous, rapid steady-state turnover, a slippery surface with an exemplary array of the asymmetric structures designed in this study (Fig. 4c and Supplementary Video 6) shows a volume of water turnover an order of magnitude greater than that of the flat slippery surfaces and other state-of-the-art surfaces (for which  $t_{\text{first}} > 200$  min as shown in Fig. 4a and Supplementary Fig. 7) developed for dropwise condensation<sup>4–7,11,12,15,17–20</sup>.

Further analysis of the steady-state turnover kinetics over 7 h (Supplementary Fig. 10) highlights that the volume of water collected on the slippery surfaces with asymmetric structures progressively outpaces that collected on the control surface, with the advantage continuously becoming greater with time even after both reach  $t_{\text{first}}$ . By coupling fast growth and fast transport, the integrated bump design not only eliminates the long onset delay observed on other surfaces but also yields a substantially higher steady-state turnover rate. This combination of short response time and reliable, high-volume long-term performance are critical in numerous applications based on condensation, transport, and phase-change heat transfer, such as heat exchange, dehumidification, and desalination systems. The central principles derived in this work further enable the manipulation of drop behaviour against an unfavourable temperature gradient. Although larger droplets are expected at a colder surface, a drop at the apex of a bump with a small radius of curvature shows the opposite trend: increased drop size at higher temperature (Supplementary Fig. 4a). This unusual behaviour is relevant to pipe geometry, an analogous form of convex curvature that is widely used in phase-change heat transfer (Supplementary Fig. 4b).

In summary, we have achieved unprecedented droplet growth and transport by designing surfaces covered with slippery asymmetric bumps based on quantitative models that integrate three mechanisms—(1) optimization of focused diffusion flux by the radius of curvature and shape of the apex, (2) asymmetric topography that guides the droplet off the bump by optimizing the free-energy profile, and (3) positive feedback between capillary transport and continued growth by coalescence along the slope.

The rapid turnover kinetics, with fast onset and sustained continuous shedding rate, combined with the ability to defy a temperature gradient, are crucial not only for water-harvesting applications, particularly in hot, arid regions where condensed water droplets will evaporate if they do not shed after a limited time, but also for many phase-change heat-transfer applications requiring reliable steady-state performance. It may also be possible to create switchable water-transporting surfaces with flexible topographical features that can be tuned by external stimuli, and to discover fundamental principles of droplet growth and transport in practical and more complex situations such as with strong airflow.

Received 13 June; accepted 15 December 2015.

Published online 24 February 2016.

1. Clus, O., Ortega, P., Muselli, M., Milimouk, I. & Beysens, D. Study of dew water collection in humid tropical islands. *J. Hydrol.* **361**, 159–171 (2008).
2. Zheng, Y. *et al.* Directional water collection on wetted spider silk. *Nature* **463**, 640–643 (2010).
3. Malik, F. T., Clement, R. M., Gethin, D. T., Krawszik, W. & Parker, A. R. Nature's moisture harvesters: a comparative review. *Bioinspir. Biomim.* **9**, 031002 (2014).
4. Miljkovic, N. *et al.* Jumping-droplet-enhanced condensation on scalable superhydrophobic nanostructured surfaces. *Nano Lett.* **13**, 179–187 (2013).
5. Rose, J. W. Dropwise condensation theory and experiment: a review. *Proc. Inst. Mech. Eng. A* **216**, 115–128 (2002).

6. Xiao, R., Miljkovic, N., Enright, R. & Wang, E. N. Immersion condensation on oil-infused heterogeneous surfaces for enhanced heat transfer. *Sci. Rep.* **3**, 1988 (2013).
7. Anand, S., Paxson, A. T., Dhiman, R., Smith, J. D. & Varanasi, K. K. Enhanced condensation on lubricant-impregnated nanotextured surfaces. *ACS Nano* **6**, 10122–10129 (2012).
8. Daniel, S., Chaudhury, M. K. & Chen, J. C. Fast drop movements resulting from the phase change on a gradient surface. *Science* **291**, 633–636 (2001).
9. Pérez-Lombard, L., Ortiz, J. & Pout, C. A review on buildings energy consumption information. *Energy Build.* **40**, 394–398 (2008).
10. Rykaczewski, K. *et al.* Dropwise condensation of low surface tension fluids on omniphobic surfaces. *Sci. Rep.* **4**, 4158 (2014).
11. Kim, P. *et al.* Liquid-infused nanostructured surfaces with extreme anti-ice and anti-frost performance. *ACS Nano* **6**, 6569–6577 (2012).
12. Quéré, D. Wetting and roughness. *Annu. Rev. Mater. Res.* **38**, 71–99 (2008).
13. Mishchenko, L., Khan, M., Aizenberg, J. & Hattori, B. D. Spatial control of condensation and freezing on superhydrophobic surfaces with hydrophilic patches. *Adv. Funct. Mater.* **23**, 4577–4584 (2013).
14. Varanasi, K. K., Hsu, M., Bhate, N., Yang, W. & Deng, T. Spatial control in the heterogeneous nucleation of water. *Appl. Phys. Lett.* **95**, 094101 (2009).
15. Nosonovsky, M. & Bhushan, B. Superhydrophobic surfaces and emerging applications: non-adhesion, energy, green engineering. *Curr. Opin. Colloid Interf. Sci.* **14**, 270–280 (2009).
16. Courbin, L. *et al.* Imbibition by polygonal spreading on microdecorated surfaces. *Nature Mater.* **6**, 661–664 (2007).
17. Wong, T.-S. *et al.* Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity. *Nature* **477**, 443–447 (2011).
18. Carlson, A., Kim, P., Amberg, G. & Stone, H. A. Short and long time drop dynamics on lubricated substrates. *Europhys. Lett.* **104**, 34008 (2013).
19. Krupenkin, T. N., Taylor, J. A., Schneider, T. M. & Yang, S. From rolling ball to complete wetting: the dynamic tuning of liquids on nanostructured surfaces. *Langmuir* **20**, 3824–3827 (2004).
20. Medici, M.-G., Mongruel, A., Royon, L. & Beysens, D. Edge effects on water droplet condensation. *Phys. Rev. E* **90**, 062403 (2014).
21. Fratzl, P. & Weinkamer, R. Nature's hierarchical materials. *Prog. Mater. Sci.* **52**, 1263–1334 (2007).
22. Parker, A. R. & Lawrence, C. R. Water capture by a desert beetle. *Nature* **414**, 33–34 (2001).
23. Nørgaard, T. & Dacke, M. Fog-basking behaviour and water collection efficiency in Namib Desert Darkling beetles. *Front. Zool.* **7**, 23 (2010).
24. Guadarrama-Cetina, J. *et al.* Dew condensation on desert beetle skin. *Eur. Phys. J. E* **37**, 109 (2014).
25. Ju, J. *et al.* A multi-structural and multi-functional integrated fog collection system in cactus. *Nature Commun.* **3**, 1247 (2012).
26. Bohn, H. F. & Federle, W. Insect aquaplaning: *Nepenthes* pitcher plants capture prey with the peristome, a fully wettable water-lubricated anisotropic surface. *Proc. Natl Acad. Sci. USA* **101**, 14138–14143 (2004).
27. Viovy, J. L., Beysens, D. & Knobler, C. M. Scaling description for the growth of condensation patterns on surfaces. *Phys. Rev. A* **37**, 4965–4970 (1988).
28. Beysens, D. Dew nucleation and growth. *C. R. Phys.* **7**, 1082–1100 (2006).
29. Zhai, L. *et al.* Patterned superhydrophobic surfaces: toward a synthetic mimic of the Namib Desert beetle. *Nano Lett.* **6**, 1213–1217 (2006).
30. Qian, M. & Ma, J. Heterogeneous nucleation on convex spherical substrate surfaces: a rigorous thermodynamic formulation of Fletcher's classical model and the new perspectives derived. *J. Chem. Phys.* **130**, 214709 (2009).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank M. Khan, J. Alvarenga, D. Daniel, S. H. Kang, M. Hoang, and J. Timonen for discussions and technical assistance. This research was supported by the Department of Energy/ARPA-E award number DE-AR0000326. N.H. thanks the Research Experiences for Undergraduates programme supported by the National Science Foundation award number DMR-1420570.

**Author Contributions** K.-C.P., P.K. and J.A. conceived the research. J.A. supervised the research. K.-C.P., P.K. and J.A. designed the slippery asymmetric bumps and the experiments. K.-C.P., P.K., N.H., D.F. and J.C.W. carried out the experiments. All authors analysed data. K.-C.P. built the analytical and numerical models. K.-C.P., P.K., A.G., D.F. and J.A. interpreted data and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.-C.P. (kpark@seas.harvard.edu) or J.A. (jaiz@seas.harvard.edu).

# Stable amorphous georgeite as a precursor to a high-activity catalyst

Simon A. Kondrat<sup>1</sup>, Paul J. Smith<sup>1</sup>, Peter P. Wells<sup>2,3</sup>, Philip A. Chater<sup>4,5</sup>, James H. Carter<sup>1</sup>, David J. Morgan<sup>1</sup>, Elisabetta M. Fiordaliso<sup>6</sup>, Jakob B. Wagner<sup>6</sup>, Thomas E. Davies<sup>1,5</sup>, Li Lu<sup>7</sup>, Jonathan K. Bartley<sup>1</sup>, Stuart H. Taylor<sup>1</sup>, Michael S. Spencer<sup>1</sup>, Christopher J. Kiely<sup>7</sup>, Gordon J. Kelly<sup>8</sup>, Colin W. Park<sup>8</sup>, Matthew J. Rosseinsky<sup>5</sup> & Graham J. Hutchings<sup>1</sup>

Copper and zinc form an important group of hydroxycarbonate minerals that include zincian malachite, aurichalcite, rosasite and the exceptionally rare and unstable—and hence little known and largely ignored<sup>1</sup>—georgeite. The first three of these minerals are widely used as catalyst precursors<sup>2–4</sup> for the industrially important methanol-synthesis and low-temperature water–gas shift (LTS) reactions<sup>5–7</sup>, with the choice of precursor phase strongly influencing the activity of the final catalyst. The preferred phase<sup>2,3,8–10</sup> is usually zincian malachite. This is prepared by a co-precipitation method that involves the transient formation of georgeite<sup>11</sup>; with few exceptions<sup>12</sup> it uses sodium carbonate as the carbonate source, but this also introduces sodium ions—a potential catalyst poison. Here we show that supercritical antisolvent (SAS) precipitation using carbon dioxide (refs 13, 14), a process that exploits the high diffusion rates and solvation power of supercritical carbon dioxide to rapidly expand and supersaturate solutions, can be used to prepare copper/zinc hydroxycarbonate precursors with low sodium content. These include stable georgeite, which we find to be a precursor to highly active methanol-synthesis and superior LTS catalysts. Our findings highlight the value of advanced synthesis methods in accessing unusual mineral phases, and show that there is room for exploring improvements to established industrial catalysts.

For the SAS precipitation process, we initially used copper(II) acetate solutions (Extended Data Fig. 1) with ethanol as the solvent; the result was amorphous copper(II) acetate, which we characterized by X-ray diffraction (XRD) and Fourier transform-infrared (FT-IR) spectroscopy. Adding 5 vol% water as a co-solvent produced a blue precipitate with an IR spectrum exhibiting a broad OH band at 3,408 cm<sup>-1</sup> and CO<sub>3</sub><sup>2-</sup> bands at 1,470, 1,404 and 829 cm<sup>-1</sup>, which are inconsistent with malachite but closely match the spectrum identifying the rare georgeite phase<sup>1</sup>. Helium pycnometry gave a density of 3.1 g cm<sup>-3</sup> for the SAS-prepared phase; as with mineralogical georgeite, the SAS-prepared phase was amorphous by XRD. We further studied the precipitate by inductively coupled plasma mass spectrometry (ICP-MS) and carbon–hydrogen–nitrogen (CHN) analysis (Extended Data Fig. 2), which gave an ideal formula for the precipitate of Cu<sub>7</sub>(CO<sub>3</sub>)<sub>5</sub>(OH)<sub>4</sub>·5H<sub>2</sub>O. We regard this to be a close match to the Cu<sub>5</sub>(CO<sub>3</sub>)<sub>3</sub>(OH)<sub>4</sub>·6H<sub>2</sub>O formula reported for georgeite<sup>1</sup>, taking into account that slight variations in composition are anticipated for an amorphous phase<sup>15</sup>. The SAS-prepared and mineralogical forms of georgeite both have a higher CO<sub>3</sub><sup>2-</sup>/Cu<sup>2+</sup> molar ratio than malachite (the CO<sub>3</sub><sup>2-</sup>/Cu<sup>2+</sup> values being 0.7, 0.6 and 0.5, respectively); this contrasts with previous assertions that georgeite and malachite (Cu<sub>2</sub>CO<sub>3</sub>(OH)<sub>2</sub>) are iso-compositional<sup>16</sup> because georgeite rapidly transformed into malachite. We did not observe such instability for SAS-prepared georgeite, as the rapid solvent extraction produced

georgeite that is effectively dry from the point of precipitation and thus has limited contact with water (which facilitates ageing and transformation to malachite).

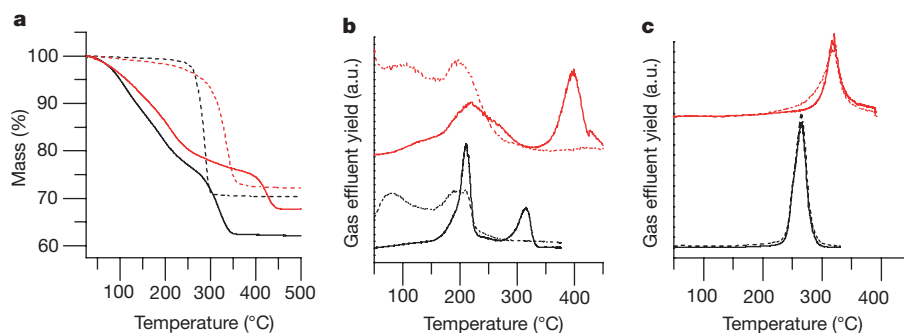
When we used mixed copper and zinc acetate solutions (with a 2/1 or 1/1 copper/zinc ratio) for the SAS precipitation, zincian georgeite (with a 2/1 or 1/1 copper/zinc ratio, as judged by ICP-MS) formed with a density, IR spectrum and XRD pattern nearly identical to those of copper-only georgeite (Extended Data Figs 1 and 2); no additional phases—such as smithsonite (ZnCO<sub>3</sub>)—were discernible by XRD. Scanning transmission electron microscopy (STEM) analysis of zincian georgeite revealed predominantly amorphous material, although a small fraction (<10 vol%) of CuO<sub>x</sub> crystallites with diameters less than 2 nanometres were observed (Extended Data Fig. 3; analysis of georgeite itself was not possible because of its instability under vacuum). In addition, the SAS-derived zincian georgeite had few impurities (such as sodium ions; Extended Data Fig. 2).

We therefore found that it was possible to synthesize stable zincian georgeite by SAS in sufficient quantities to evaluate the catalytic performance of copper/zinc oxide, derived from zincian georgeite, in the methanol-synthesis and LTS reactions. As a control, we synthesized a zincian-malachite precursor—considered the optimal precursor for model copper/zinc-oxide catalysts<sup>3</sup>—by co-precipitation (see Methods). The catalysts used here were not stabilized with alumina, an important promoter and stabilizer of the commercial catalyst<sup>17,18</sup>. Before use, the precursors were subjected to calcination to remove the carbonate, followed by *in situ* reduction to form the active catalyst. The optimal calcination temperature was determined using thermal gravimetric analysis (TGA) with evolved gas analysis (EGA), which showed SAS-prepared georgeite to exhibit three distinct mass losses, and malachite to exhibit a single mass loss, associated with concurrent removal of water and carbon dioxide (Fig. 1). The multistep decomposition of georgeite and zincian georgeite (the 1/1 copper/zinc sample shown in Extended Data Fig. 2) can be separated into three regions: water loss at 80–100 °C, water and carbon dioxide loss at 190–210 °C, and carbon dioxide loss at 315–420 °C (carbonate decomposition). On the basis of this analysis, the precursors were calcined at 300 °C in air for 6 hours: this is below the temperature associated with the evolution of carbon dioxide, and residual carbonate species have been shown to improve the activity of methanol-synthesis catalysts<sup>19</sup>. The calcined material was then reduced *in situ* in dilute hydrogen at 225 °C.

Figure 2 contrasts the activity of our model catalysts with that of commercial catalysts, showing that the catalyst derived from SAS-prepared zincian georgeite is more active than both the catalyst formed from zincian malachite and the standard commercial catalyst. In the case of methanol synthesis, for which our model catalysts showed markedly higher initial productivity (normalized for copper surface area;

<sup>1</sup>Cardiff Catalysis Institute, School of Chemistry, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, UK. <sup>2</sup>The UK Catalysis Hub, Research Complex at Harwell, Harwell, Oxon OX11 0FA, UK. <sup>3</sup>Kathleen Lonsdale Building, Department of Chemistry, University College London, Gordon Street, London WC1H 0AJ, UK. <sup>4</sup>Diamond Light Source, Didcot OX11 0DE, UK. <sup>5</sup>Department of Chemistry, University of Liverpool, Crown Street, Liverpool L69 7ZD, UK. <sup>6</sup>Center for Electron Nanoscopy, Technical University of Denmark, Fysikvej 307, DK-2800 Kgs Lyngby, Denmark. <sup>7</sup>Department of Materials Science and Engineering, Lehigh University, 5 East Packer Avenue, Bethlehem, Pennsylvania 18015, USA. <sup>8</sup>Johnson Matthey, PO Box 1, Belasis Avenue, Cleveland TS23 1LB, UK.





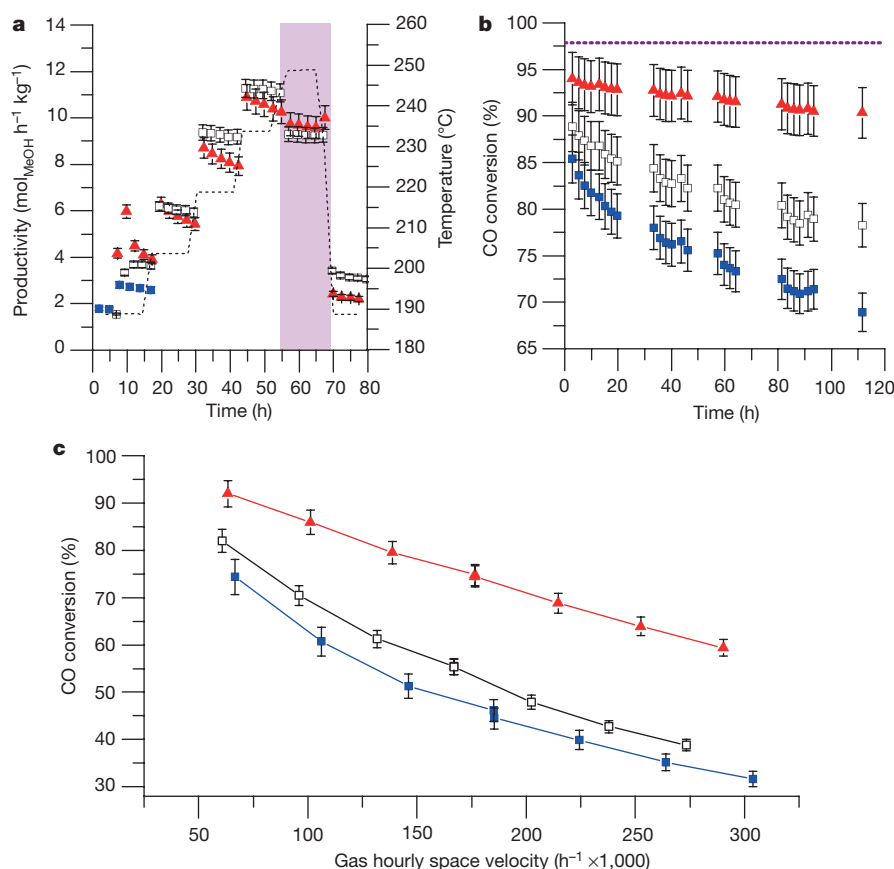
**Figure 1 | Thermal gravimetric analysis (TGA) with evolved gas analysis (EGA) of georgeite and malachite.** **a**, TGA of georgeite (black solid line), zincian georgeite (red solid line), malachite (black dotted line) and zincian malachite (red dotted line). **b**, EGA of georgeite (black lines) and

zincian georgeite (red lines). **c**, EGA of malachite (black lines) and zincian malachite (red lines). In **b** and **c**, the solid lines indicate  $\text{CO}_2$  (fragment with mass 44) and dashed lines indicate  $\text{H}_2\text{O}$  (fragment with mass 18).

Extended Data Fig. 4), the enhancement of activity was short-lived. But in the LTS reaction, the SAS-prepared zincian-georgeite-derived catalyst was far more active and stable over the entire test period; indeed, enhanced activity was apparent at all contact times investigated (Fig. 2). These observations demonstrate that the availability of stable zincian georgeite as a precursor gives access to highly active copper/zinc-oxide catalysts, with the LTS activity and stability maintained

using 30% less copper than is present in a catalyst derived from a 2/1 copper/zinc georgeite (Extended Data Fig. 4).

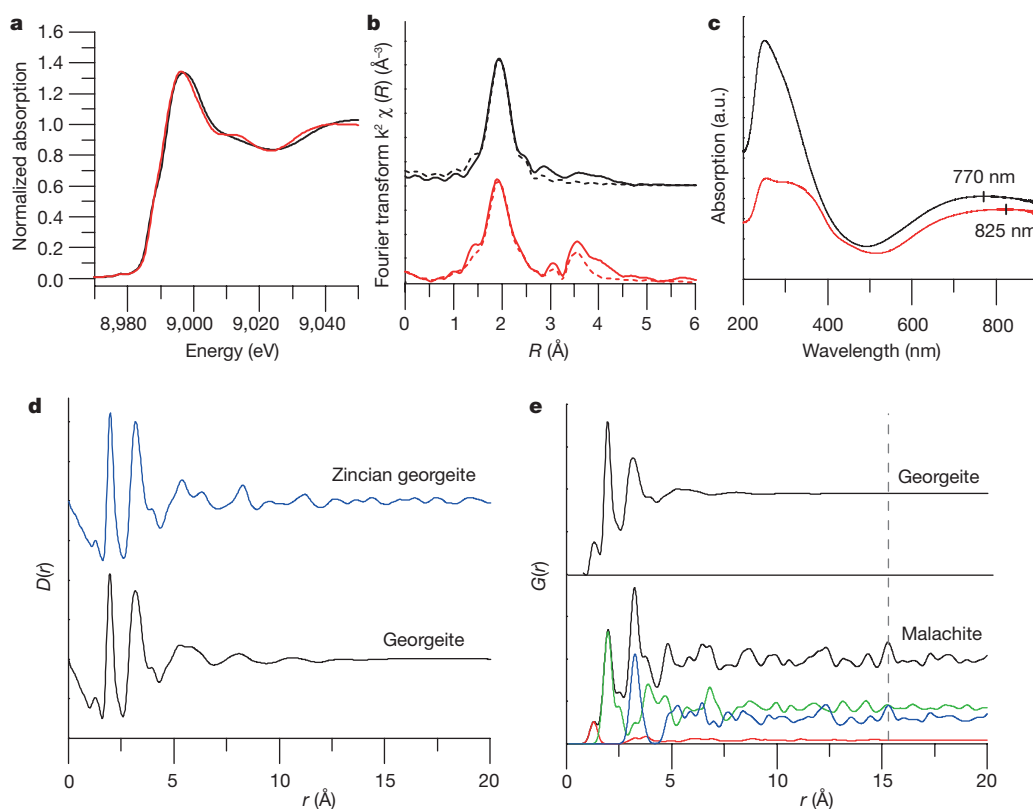
The stability and quantities of georgeite that are available through SAS allowed us to explore structural origins for its enhanced activity. X-ray-absorption near-edge structure (XANES) data for copper-only georgeite and malachite precursors (Fig. 3a) indicate that both materials contain  $\text{Cu}^{2+}$  in a distorted octahedral environment<sup>20</sup>, in agreement



**Figure 2 | Comparison of our model catalysts to commercial catalysts.**

**a**, Methanol synthesis (shown as mass-normalized, time-on-line methanol production). The dashed line denotes the representative reactor-bed temperature for methanol synthesis. Reaction conditions for methanol synthesis: 190–250 °C (shaded area shows reaction under equilibrium conditions at 250 °C), 25 bar, gas composition  $\text{CO}/\text{CO}_2/\text{H}_2/\text{N}_2 = 6/9.2/67/17.8$ , mass hourly space velocity (MHSV) =  $7,200 \text{ kg}^{-1} \text{ h}^{-1}$ . The selectivities of all of the catalysts to methanol are greater than 99.96%. **b**, Low-temperature water-gas shift (LTS) reaction (shown as time-on-line CO conversion). The dashed line represents the maximum equilibrium conversion. Reaction conditions for LTS: 220 °C,

27.5 bar, gas composition  $\text{H}_2\text{O}/\text{CO}/\text{CO}_2/\text{H}_2/\text{N}_2 = 50/2/8/27.5/12.5$ , MHSV =  $75,000 \text{ kg}^{-1} \text{ h}^{-1}$ . **c**, LTS conversions at different gas hourly space velocities. (Extended Data Fig. 4 shows LTS mass-normalized data, copper mass-normalized productivities/activities, initial copper surface area normalized productivities/activities and product impurities.) Red triangles, zincian-georgeite-derived catalyst; blue squares, co-precipitated zincian-malachite-derived catalyst; open squares, industrial standards (composition of the industrial methanol-synthesis catalyst by wt% is  $\text{CuO}/\text{ZnO}/\text{Al}_2\text{O}_3 = 60/30/10$ ; composition of the industrial LTS catalyst by wt% is  $\text{CuO}/\text{ZnO}/\text{Al}_2\text{O}_3 = 50/33/17$ ). Error-bar calculations are discussed in Methods.



**Figure 3 | X-ray absorption fine-edge spectroscopy (XAFS) and X-ray pair distribution function (PDF) analysis.** **a**, Copper K-edge XAFS information on local structure, for malachite (red line) and georgeite (black line). **b**, Fourier transform (FT) of extended X-ray absorption fine-edge spectra (EXAFS)  $k^2$  weighted  $\chi$  data (Fourier transform  $k^2 \chi(R)$ ), with associated fitting parameters for malachite (red line) and georgeite (black line). Variation in the magnitude of the FT is plotted with distance  $R$  (Å) from the Cu absorber;  $k$  denotes the fitting window from the  $\chi$  data. The fit (shown with dashed lines) was modelled on four shorter Cu–O bond distances and two longer Jahn–Teller distorted Cu–O bond distances (Extended Data Table 1). **c**, Ultraviolet–visible spectra for malachite (red line) and georgeite (black line), with centres of d–d

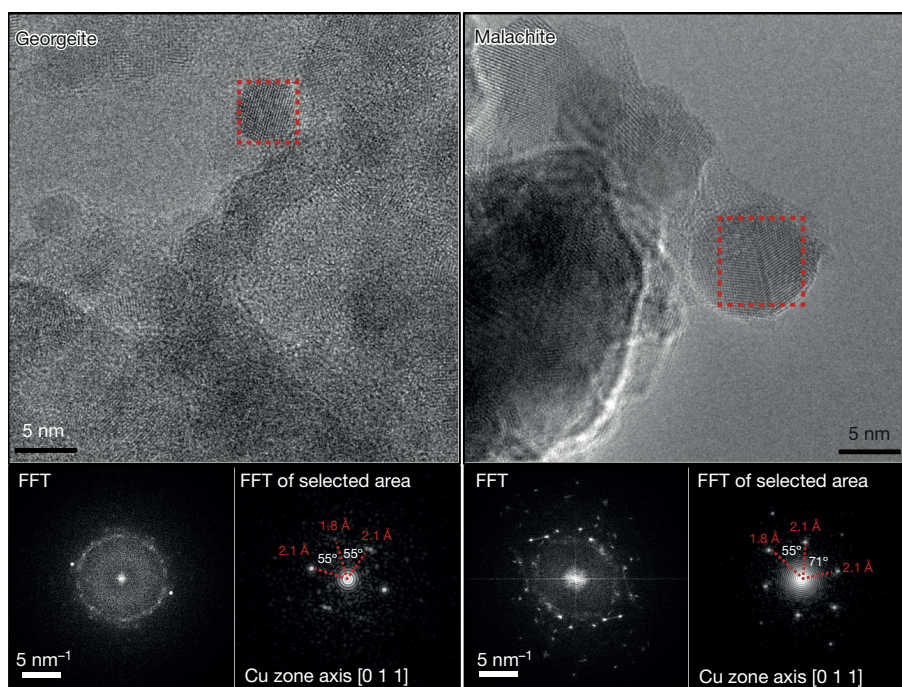
transition bands (based on peak maxima) marked as small vertical lines (zincian phases are shown in Extended Data Fig. 5a). **d**, X-ray PDF data for georgeite and zincian georgeite.  $D(r)$  is the probability of finding a pair of atoms, weighted by the scattering power of the pair, at a given distance,  $r$ . The probability is scaled to  $r$  to emphasize order at greater distances. **e**, Top, observed PDF  $G(r)$  data ( $G(r)$  is equivalent to  $D(r)$ , only without scaling by  $r$ ) for georgeite. Bottom, calculated PDF  $G(r)$  data for all of the atom pairs of malachite (black), together with the partial contributions from the most strongly contributing atom pairs: C–O (red), Cu–O (green) and Cu–Cu (blue). The dotted line represents a contribution from crystallographically equivalent Cu atoms in malachite (see text).

with diffuse reflectance ultraviolet–visible spectroscopy (DRS) and extended X-ray absorption fine structure (EXAFS) data (Fig. 3b and Extended Data Table 1). But our analysis also suggests subtle differences in coordination geometries. Specifically, DRS locates the optical band associated with the metal centre(s) in georgeite at 770 nm, whereas the combined bands from the two crystallographic copper centres in malachite occur at 825 nm (ref. 21). Although XANES data for georgeite and malachite are similar, the Fourier transform of the EXAFS  $\chi$  data (Fig. 3b) shows that malachite has contributions further out from the copper absorber that are correlated to the scattering effects of copper–copper neighbours, whereas there are very few features in the georgeite EXAFS data beyond 2 Å and no evidence of metal–metal correlation. The addition of zinc did not affect metal geometry in georgeite, with no discernible changes appearing in the copper EXAFS data or zinc K-edge XANES or EXAFS  $\chi$ -plots of zincian georgeite (such changes would be expected in the case of impurity phases such as  $\text{ZnCO}_3$ ) (Extended Data Fig. 5).

X-ray pair distribution function (PDF) analysis, which does not suffer from phenomena related to scattering-path phasing, indicated that georgeite and zincian georgeite have strikingly similar and strong short-range order up to 5 Å (Fig. 3d); subtle differences appear between 5 Å and 10 Å, with georgeite exhibiting less long-range order. The extended order seen upon addition of zinc could be responsible for the enhanced stability of zincian georgeite prepared by co-precipitation<sup>16</sup>. A comparison of the georgeite PDF with a calculated PDF of malachite<sup>22</sup>

(Fig. 3e) shows that georgeite is not simply a nanoscale form of malachite (a phenomenon noted for other amorphous minerals, such as iron sulphide<sup>23</sup>): the complete absence of the 15.27 Å correlation between crystallographically identical copper atoms in georgeite shows that local ordering operates well below the length scales associated with the malachite unit cell. PDF matching to other copper phases such as aurichalcite<sup>24</sup>, azurite<sup>25</sup> and rosinite<sup>26</sup> confirmed georgeite to be a material distinct from any known copper hydroxycarbonate (Extended Data Fig. 5).

After synthesis, the zincian georgeite and malachite precursors were calcined at 300 °C to form copper-oxide/zinc-oxide intermediates, and then reduced to the active catalyst phase. We find that the intermediate materials produced on calcination and the final-state catalysts both exhibit a microstructure that reflects the structural characteristics of their precursor phase. In the case of SAS-prepared zincian georgeite, calcination produced disordered copper oxide/zinc oxide with nanocrystals of 3–4 nm in diameter, as observed by STEM and PDF analysis (Extended Data Figs 6 and 7). The extent of the disorder resulted in XRD and XANES analysis showing no discernible metal-oxide contribution, although subtle changes in the EXAFS data were observed (Extended Data Fig. 8). Disordered copper-oxide species have previously been reported to have copper-edge XANES data that fit with a Jahn–Teller distorted octahedron<sup>27</sup>, as with the calcined zincian georgeite. We found that the nanocrystallinity of the SAS-prepared zincian georgeite persisted until the residual high-temperature carbonate



**Figure 4 | Characterization, by environmental transmission electron microscopy, of the microstructure of the reduced final-state catalysts derived from zincian georgeite and zincian malachite.** Calcined materials were reduced under a  $\text{H}_2$  pressure of 2 mbar and a temperature of 225 °C. The representative images shown in the top panels reveal distinct Cu nanoparticles distributed on ZnO for both samples. This was confirmed by fast Fourier transform (FFT) analysis, shown in the bottom panels. The reduced malachite sample generally has larger Cu nanoparticles than the georgeite sample. Also, there is a substantially

decomposed to form XRD-discernible copper-oxide and zinc-oxide phases above 425 °C (Extended Data Fig. 7). Conversely, calcination of zincian malachite at temperatures greater than 300 °C produced crystalline, XRD- and XANES-observable, copper-oxide and zinc-oxide phases (Extended Data Fig. 7). A linear combination fit of the XANES data showed that 60% of the 300 °C calcined zincian malachite (Extended Data Fig. 8) was associated with copper oxide and zinc oxide, whereas these phases were not present in the calcined zincian georgeite. STEM showed that 300 °C calcined zincian malachite (Extended Data Fig. 6) was composed of copper-oxide and zinc-oxide grains with much longer-range crystalline order than the corresponding calcined zincian georgeite.

The nanocrystalline nature of calcined zincian georgeite translates to a high copper surface area ( $53 \pm 3 \text{ m}^2 \text{ g}^{-1}$ ) and small mean copper crystallite size ( $7.4 \pm 0.7 \text{ nm}$ ) when the material is reduced. Environmental transmission electron microscopy (ETEM) carried out under reducing conditions shows a complex mixture of copper particles and exceptionally small, disordered zinc-oxide grains (the mean zinc-oxide crystallite size was  $2.8 \pm 1.6 \text{ nm}$ , as judged by XRD), with notable interactions between copper and zinc oxide (Fig. 4). By comparison, the more crystalline zincian-malachite-derived copper oxide/zinc oxide had a lower copper surface area ( $35 \text{ m}^2 \text{ g}^{-1}$ ) and a microstructure composed of distinctly larger copper particles (mean crystallite size was  $14.2 \pm 3.8 \text{ nm}$ , by XRD) and zinc oxide particles (mean crystallite size was  $5.6 \pm 3.6 \text{ nm}$  by XRD). This difference in particle size and structural order is reflected in the smaller copper–copper and zinc–zinc coordination numbers noted for the zincian-georgeite-derived catalyst from *in situ* EXAFS analysis (Extended Data Fig. 9). On a macroscopic level, strong interactions between copper and zinc oxide, as shown by XPS observations of partial  $\text{ZnO}_x$  coverage of copper on reduction<sup>8</sup>, were observed for both catalysts (Extended Data Fig. 9). However, ETEM shows considerably more interactions between

greater Cu–ZnO interaction in the georgeite-derived sample than in the malachite-derived catalyst. The georgeite sample has a complex mix of Cu and poorly defined ZnO particles (as revealed by diffuse rings in the large-scale FFTs); the malachite sample shows better-defined, phase-separated Cu and ZnO. The oxidation state of Cu was monitored by means of electron energy-loss spectroscopy, which revealed that Cu was in its metallic state in both samples under the reduction conditions (Extended Data Fig. 9f). The Cu zone axis is the direction (as determined by analysis of the FFT) along which the Cu crystallite is being observed.

copper and zinc oxide in the catalyst that was derived from disordered zincian georgeite.

Methanol-synthesis activity is well known to correlate strongly with copper surface area<sup>28</sup>, and a large interfacial area between copper and zinc oxide has been reported to produce highly active methanol-synthesis catalysts<sup>29</sup>, containing active sites that are associated with surface-defect copper sites decorated by zinc<sup>8</sup>. However, we find that 40 hours of LTS conditions markedly decrease the copper surface areas of catalysts derived from both zincian georgeite ( $53 \pm 3 \text{ m}^2 \text{ g}^{-1}$  to  $17 \pm 1 \text{ m}^2 \text{ g}^{-1}$ ) and zincian malachite ( $38 \pm 2 \text{ m}^2 \text{ g}^{-1}$  to  $19 \pm 1 \text{ m}^2 \text{ g}^{-1}$ )—yet the catalysts were still active after this time period, clearly indicating that copper surface areas do not adequately correlate with catalyst activity in this reaction. We surmise that the high activity of our catalyst can be explained by a combination of a greater content of exposed copper, and an intimate interface contact between copper and zinc oxide, inherited from the zincian-georgeite precursor. Moreover, the SAS method produces very pure zincian georgeite with a low content of sodium ions, without requiring an aqueous washing step; this high purity might also contribute to both the high activity and the stability of the zincian-georgeite-derived LTS catalyst.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 11 September; accepted 8 December 2015.**

**Published online 15 February 2016.**

1. Bridge, P. J., Just, J. & Hey, M. H. Georgeite, a new amorphous copper carbonate from the Carr Boyd Mine, Western Australia. *Mineral. Mag.* **43**, 97–98 (1979).
2. Spencer, M. S. The role of zinc oxide in Cu/ZnO catalysts for methanol synthesis and the water-gas shift reaction. *Top. Catal.* **8**, 259–266 (1999).
3. Behrens, M. Meso- and nano-structuring of industrial Cu/ZnO/( $\text{Al}_2\text{O}_3$ ) catalysts. *J. Catal.* **267**, 24–29 (2009).
4. Fujitani, T. & Nakamura, J. The effect of ZnO in methanol synthesis catalysts on Cu dispersion and the specific activity. *Catal. Lett.* **56**, 119–124 (1998).



5. Chinchin, G. C., Denny, P. J., Jennings, J. R., Spencer, M. S. & Waugh, K. C. Synthesis of methanol. Part 1. Catalysts and kinetics. *Appl. Catal.* **36**, 1–65 (1988).
6. Rhodes, C., Hutchings, G. J. & Ward, A. M. Water-gas shift reaction: finding the mechanistic boundary. *Catal. Today* **23**, 43–58 (1995).
7. Short, G. D., Chinchin, G. C. & Williamson, J. G. Synthesis of methanol; finely divided oxides of zinc, aluminum, magnesium; with metallic copper. US patent 4,788,175 (1988).
8. Behrens, M. *et al.* The active site of methanol synthesis over Cu/ZnO/Al<sub>2</sub>O<sub>3</sub> industrial catalysts. *Science* **336**, 893–897 (2012).
9. Baltes, C., Vukojevic, S. & Schueth, F. Correlations between synthesis, precursor, and catalyst structure and activity of a large set of CuO/ZnO/Al<sub>2</sub>O<sub>3</sub> catalysts for methanol synthesis. *J. Catal.* **258**, 334–344 (2008).
10. Bems, B. *et al.* Relations between synthesis and microstructural properties of copper/zinc hydroxycarbonates. *Chemistry* **9**, 2039–2052 (2003).
11. Pollard, A. M. *et al.* Georgeite and azurite as precursors in the preparation of coprecipitated copper/zinc oxide catalysts. *Appl. Catal. A* **85**, 1–11 (1992).
12. Prieto, G., de Jong, K. P. & de Jongh, P. E. Towards 'greener' catalyst manufacture: reduction of wastewater from the preparation of Cu/ZnO/Al<sub>2</sub>O<sub>3</sub> methanol synthesis catalysts. *Catal. Today* **215**, 142–151 (2013).
13. Tang, Z.-R. *et al.* New nanocrystalline Cu/MnO<sub>x</sub> catalysts prepared from supercritical antisolvent precipitation. *ChemCatChem* **1**, 247–251 (2009).
14. Reverchon, E. Supercritical antisolvent precipitation of micro- and nano-particles. *J. Supercrit. Fluids* **15**, 1–21 (1999).
15. Rogers, A. P. A review of the amorphous minerals. *J. Geol.* **25**, 515–541 (1917).
16. Pollard, A. M., Thomas, R. G., Williams, P. A., Just, J. & Bridge, P. J. The synthesis and composition of georgeite and its reactions to form other secondary copper(II) carbonates. *Mineral. Mag.* **55**, 163–166 (1991).
17. Behrens, M. *et al.* Performance improvement of nanocatalysts by promoter-induced defects in the support material: methanol synthesis over Cu/ZnO:Al. *J. Am. Chem. Soc.* **135**, 6061–6068 (2013).
18. Campbell, J. S. Influences of catalyst formulation and poisoning on the activity and die-off of low temperature shift catalysts. *Ind. Eng. Chem. Process Des. Dev.* **9**, 588–595 (1970).
19. Schur, M. *et al.* Continuous coprecipitation of catalysts in a micromixer: nanostructured Cu/ZnO composite for the synthesis of methanol. *Angew. Chem. Int. Edn* **42**, 3815–3817 (2003).
20. Rothe, J., Hormes, J., Bonnemann, H., Brijoux, W. & Siepen, K. In situ X-ray absorption spectroscopy investigation during the formation of colloidal copper. *J. Am. Chem. Soc.* **120**, 6019–6023 (1998).
21. Klokishner, S. *et al.* Cation ordering in natural and synthetic (Cu<sub>1-x</sub>Zn<sub>x</sub>)<sub>2</sub>CO<sub>3</sub>(OH)<sub>2</sub> and (Cu<sub>1-x</sub>Zn<sub>x</sub>)<sub>5</sub>(CO<sub>3</sub>)<sub>2</sub>(OH)<sub>6</sub>. *J. Phys. Chem. A* **115**, 9954–9968 (2011).
22. Behrens, M. & Girdsies, F. Structural effects of Cu/Zn substitution in the malachite-rosasite System. *Z. Anorg. Allg. Chem.* **636**, 919–927 (2010).
23. Michel, F. M. *et al.* Short- to medium-range atomic order and crystallite size of the initial FeS precipitate from pair distribution function analysis. *Chem. Mater.* **17**, 6246–6255 (2005).
24. Harding, M. M., Kariuki, B. M., Cernik, R. & Cressey, G. The structure of aurichalcite, (CuZn)<sub>5</sub>(OH)<sub>6</sub>(CO<sub>3</sub>)<sub>2</sub> determined from a microcrystal. *Acta Crystallogr. B* **50**, 673–676 (1994).
25. Belokoneva, E. L., Gubina, Y. K. & Forsyth, J. B. The charge density distribution and antiferromagnetic properties of azurite Cu<sub>3</sub>[CO<sub>3</sub>]<sub>2</sub>(OH)<sub>2</sub>. *Phys. Chem. Miner.* **28**, 498–507 (2001).
26. Perchiazzi, N. & Merlino, S. The malachite-rosasite group: crystal structures of glaukosphaerite and pokrovskite. *Eur. J. Mineral.* **18**, 787–792 (2006).
27. Belin, S. *et al.* CuAu/SiO<sub>2</sub> catalysts for the selective oxidation of propene to acrolein: the impact of catalyst preparation variables on material structure and catalytic performance. *Catal. Sci. Technol.* **3**, 2944–2957 (2013).
28. Chinchin, G. C., Waugh, K. C. & Whan, D. A. The activity and state of the copper surface in methanol synthesis catalysts. *Appl. Catal.* **25**, 101–107 (1986).
29. Behrens, M. *et al.* The potential of microstructural optimization in metal/oxide catalysts: higher intrinsic activity of copper by partial embedding of copper nanoparticles. *Chem. Cat. Chem.* **2**, 816–818 (2010).

**Acknowledgements** We thank C. Brookes, L. Van de Water, H. Stanness and C. Ramson for technical assistance. We thank Johnson Matthey and the Engineering and Physical Sciences Research Council (EPSRC) for funding through a CASE award, and acknowledge funding from the UK Technology Strategy Board and the EPSRC and UK Catalysis Hub (grants EP/K014714/1, EP/K014714/1, EP/K014668/1, EP/K014706/1, EP/H000925/1 and EP/I019693/1). We used the B18 beamline at the Diamond Light Source (allocation number SP8071) with the help of D. Gianolio and G. Cibir; and we used the I15 beamline, which contributed to the PDF results. This research used the resources of the Advanced Photon Source (APS)—a US Department of Energy (DOE) Office of Science User Facility, operated for the DOE Office of Science by Argonne National Laboratory under contract DE-AC02-06CH11357; we thank K. Chapman for collecting PDF data at beamline 11-ID-B, APS. C.J.K. acknowledges funding from the National Science Foundation Major Research Instrumentation program (GR#MRI/DMR-1040229). M.J.R. is a Royal Society Research Professor. We thank the A.P. Møller and Chastine Mc-Kinney Møller Foundation for contributing to the establishment of the Center for Electron Nanoscopy in the Technical University of Denmark. Source data for Figs 1–3 are available at <http://dx.doi.org/10.17035/d.2015.0008102108>.

**Author Contributions** S.A.K., J.K.B., S.H.T., M.S.S., G.J.K., C.W.P., M.J.R., C.J.K. and G.J.H. designed the experiments. S.A.K. and P.J.S. prepared samples for TGA/EGA analysis and IR spectroscopy; S.A.K., P.J.S. and J.H.C. carried out the catalysis evaluation and determination of copper surface area; S.A.K., P.J.S. and P.A.C. carried out the XRD analysis; P.A.C. carried out the PDF analysis; P.P.W., S.A.K. and P.J.S. carried out the XAFS and interpretation; L.L. and C.J.K. carried out the TEM and interpretation; E.M.F., J.B.W., C.J.K. and S.A.K. carried out the ETEM and interpretation; D.J.M., P.J.S. and S.A.K. carried out the XPS and interpretation. S.A.K., C.J.K., J.B.W., G.J.K., M.J.R. and G.J.H. wrote and edited the manuscript. G.J.H. directed the research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.J.H. (Hutch@cardiff.ac.uk) and S.A.K. (KondratSA@cardiff.ac.uk).

## METHODS

**Material preparation.** SAS method for preparing georgeite and zincian georgeite. Copper(II) acetate monohydrate ( $4 \text{ mg ml}^{-1}$ ) and zinc(II) acetate dihydrate ( $2.16 \text{ mg ml}^{-1}$ ) (Sigma Aldrich  $\geq 99\%$  Puriss) were dissolved in ethanol (reagent grade, Fischer Scientific) containing 0 vol%, 5 vol% or 10 vol% deionized water. Smithsonite  $\text{ZnCO}_3$  was prepared with zinc(II) acetate dehydrate ( $2.16 \text{ mg ml}^{-1}$ ) in a 10 vol% water and ethanol solution. SAS experiments were performed using apparatus manufactured by Separex.  $\text{CO}_2$  (from BOC) was pumped through the system (held at 110 bar,  $40^\circ\text{C}$ ) via the outer part of a coaxial nozzle at a rate of  $6 \text{ kg h}^{-1}$ . The metal salt solution was concurrently pumped through the inner nozzle, using an Agilent HPLC pump at a rate of  $6.4 \text{ ml min}^{-1}$ . The resulting precipitate was recovered on a stainless steel frit, while the  $\text{CO}_2$ -solvent mixture passed down stream, where the pressure was decreased to separate the solvent and  $\text{CO}_2$ . The precipitation vessel has an internal volume of 1 litre. Precipitation was carried out for 120 minutes, and followed by a purge of the system with ethanol- $\text{CO}_2$  for 15 minutes, then  $\text{CO}_2$  for 60 minutes under 110 bar and  $40^\circ\text{C}$ . The system was then depressurized and the dry powder collected. Recovered samples were placed in a vacuum oven at  $40^\circ\text{C}$  for 4 hours to remove any residual solvent. Approximately 1.5 g of georgeite is prepared during the 120-minute duration of solution precipitation. Co-precipitation method for preparing malachite and zincian malachite precursors for the standard methanol-synthesis catalysts. The procedure was performed via a semi-continuous process, using two peristaltic pumps to maintain pH. Copper(II) nitrate hydrate and zinc(II) nitrate hydrate solutions in deionized water were prepared with copper/zomc molar ratios of 1/0, 1/1 and 2/1. The premixed metal solution ( $5 \text{ l}$ ,  $\sim 0.5 \text{ g ml}^{-1}$ ) was preheated along with a separate  $5 \text{ l}$  solution of  $1.5 \text{ M}$  sodium carbonate. The mixed metals were precipitated by combining the two solutions concurrently at  $65^\circ\text{C}$ , with the pH being held between 6.5 and 6.8. The precipitate would spill over from the small precipitation pot into a stirred ageing vessel, held at  $65^\circ\text{C}$ . The precipitate was aged for 15 minutes after all precursor solutions had been used.

The precipitate was then filtered and washed to minimize sodium content. The sample was washed with  $6 \text{ l}$  of hot deionized water, and the sodium content monitored using a photometer. The washing process was repeated until the sodium content showed no change. The sample was then dried at  $110^\circ\text{C}$  for 16 hours. Samples were calcined for 6 hours at  $300^\circ\text{C}$  or  $450^\circ\text{C}$  in a tube furnace under static air. The ramp rate used to reach the desired temperature was  $1^\circ\text{C min}^{-1}$ .

**Catalyst testing.** Catalyst testing was performed with  $0.5 \text{ g}$  of the calcined catalyst, pelleted and ground to a sieve fraction of  $0.6\text{--}1 \text{ mm}$  for both the methanol-synthesis and the LTS test reactions. The catalysts were reduced *in situ* using a  $2\% \text{ H}_2/\text{N}_2$  gas mixture at  $225^\circ\text{C}$  (ramp rate  $1^\circ\text{C min}^{-1}$ ), before the reaction gases were introduced. Data reported for the industrial standards are the mean value from four repeat experiments. Error bars are based on the standard deviation from the mean values.

**Methanol synthesis.** Testing was carried out in a single-stream, six-fixed-bed reactor with an additional bypass line. After reduction, the catalysts were then subjected to synthesized syngas ( $\text{CO}/\text{CO}_2/\text{H}_2/\text{N}_2 = 6.9/2/67/17.8$ ) at  $3.51 \text{ h}^{-1}$ ,  $25 \text{ bar}$  pressure and  $195^\circ\text{C}$ . In-line gas analysis was performed using an FT-IR spectrometer, which detected  $\text{CO}$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{O}$  and  $\text{CH}_3\text{OH}$ . Downstream of the catalyst beds, knockout pots collected effluent produced from the reaction. The contents were collected after each test run and analysed using gas chromatography to evaluate the selectivity of catalysts. The total system flow was maintained using the bypass line. LTS. Testing was performed in six parallel fixed-bed reactors with a single stream feed and an additional bypass line. After reduction, the catalysts were subjected to synthetic syngas ( $\text{CO}/\text{CO}_2/\text{H}_2/\text{N}_2 = 1/4/13.75/6.25$ ) at  $27.5 \text{ bar}$  pressure and  $220^\circ\text{C}$ . The reactant gas stream was passed through vaporized water to give a water composition of  $50 \text{ vol}\%$ . This gives a total gas flow of  $\text{H}_2\text{O}/\text{CO}/\text{CO}_2/\text{H}_2/\text{N}_2 = 50/2/8/27.5/12.5$ . The standard mass hourly space velocity (MHSV) used for testing was  $75,000 \text{ h}^{-1} \text{ kg}^{-1}$ . In-line IR analysis was carried out to measure  $\text{CO}$  conversion. Selectivity was determined by the methanol content within the knockout pots downstream of the reactors. Space and mass velocity variation tests were performed at 65 hours' time-on-line by altering the flow for each catalyst bed. Relative activities were calculated by altering the flow for each catalyst bed in order to achieve  $75\%$   $\text{CO}$  conversion after 75 hours' time-on-line. The total system flow was maintained using the bypass line.

**Characterization methods.** Powder X-ray diffraction (XRD). Measurements were performed using a PANalytical X'pert Pro diffractometer with a Ni-filtered  $\text{CuK}_\alpha$  radiation source operating at  $40 \text{ kV}$  and  $40 \text{ mA}$ . Patterns were recorded over the range of  $10\text{--}80^\circ 2\theta$  using a step size of  $0.016^\circ$ . All patterns were matched using the International Centre for Diffraction Data (ICDD) database. An *in situ* Anton Parr XRK900 cell (with an internal volume of  $\sim 0.5 \text{ l}$ ) was used to monitor the formation of metallic copper during the reduction of the  $\text{CuO}/\text{ZnO}$  materials. A flow of  $2\% \text{ H}_2/\text{N}_2$  ( $60 \text{ ml min}^{-1}$ ) was passed through the sample bed while the cell was heated to  $225^\circ\text{C}$  (ambient temperature to  $125^\circ\text{C}$ , ramp rate  $10^\circ\text{C min}^{-1}$ ;  $125\text{--}225^\circ\text{C}$ , ramp

rate  $2^\circ\text{C min}^{-1}$ ). The sample was then cooled to room temperature and a  $20\text{--}100^\circ 2\theta$  scan performed. Copper crystallite size was estimated from a peak-broadening analysis of the XRD pattern using Topas Academic<sup>30</sup>, and the volume-weighted column height ( $L_{\text{vol}}$ ) was calculated according to ref. 31.

**X-ray absorption fine-edge spectroscopy (XAFS).** K-edge XAFS studies of copper and zinc were carried out on the B18 beamline at the Diamond Light Source, Didcot, UK. Measurements were performed using a quick extended (QE) XAFS set-up with a fast-scanning silicon (111) double-crystal monochromator. The time resolution of the spectra reported herein was 1 minute per spectrum ( $k_{\text{max}} = 14$ ), and on average three scans were acquired to improve the signal-to-noise ratio of the data. All *ex situ* samples were diluted with cellulose and pressed into pellets to optimize the effective-edge step of the XAFS data and measured in transmission mode using ion-chamber detectors. All transmission XAFS spectra were acquired concurrently with the appropriate reference foil (copper or zinc) placed between  $I_1$  and  $I_{\text{ref}}$  XAS data processing, and EXAFS analysis were performed using IFEFFIT<sup>32</sup> with the Horae package<sup>33</sup> (Athena and Artemis). The FT data were phase-corrected for oxygen. The amplitude-reduction factor,  $s_0^2$ , was derived from EXAFS data analysis of a known copper reference compound, namely tenorite ( $\text{CuO}$ ). For the fitting of the local coordination geometry of georgeite and malachite, the Jahn-Teller distorted copper-oxygen bond was difficult to observe because of thermal disorder, and so was fixed in the model.

**X-ray pair distribution function (PDF) analysis.** Synchrotron X-ray PDF data were collected on the 11-ID-B beamline at the Advanced Photon Source, Argonne National Laboratory, and on the I15 beamline at the Diamond Light Source. Powder samples were packed into kapton capillaries with an internal diameter of  $1 \text{ mm}$ . Room-temperature powder X-ray diffraction data were collected at a wavelength of  $0.2114 \text{ \AA}$  (11-ID-B) and  $0.1620 \text{ \AA}$  (I15) using the Rapid Acquisition PDF method<sup>34</sup>. The scattering data ( $0.5 \leq Q \leq 22 \text{ \AA}^{-1}$ ) were processed into PDF data using the program GudrunX<sup>35</sup>. Two notations of PDF data are presented in this manuscript:  $G(r)$  and  $D(r)$  (ref. 36). The total radial distribution function,  $G(r)$ , is the probability of finding a pair of atoms, weighted by the scattering power of the pair, at a given distance,  $r$ ; it is the Fourier transform experimentally determined total structure factor.  $D(r)$  is re-weighted to emphasize features at high  $r$ , such that  $D(r) = 4\pi r \rho_0 G(r)$ , where  $\rho_0$  is the average number density of the material (in atoms  $\text{\AA}^{-3}$ ).

**Fourier transform-infrared spectroscopy (FT-IR).** Analysis was carried out using a Jasco FT/IR 660 Plus spectrometer in transmission mode over a range of  $400\text{--}4,000 \text{ cm}^{-1}$ . Catalysts were supported in a pressed KBr disk.

**Raman spectroscopy.** Raman spectra were obtained using a Renishaw inVia spectrometer equipped with an argon ion laser ( $\lambda = 514 \text{ nm}$ ).

**Thermal gravimetric analysis (TGA).** TGA measurements were performed using a SETARAM Labsys analyser with sample masses of about  $20 \text{ mg}$ , at  $1^\circ\text{C min}^{-1}$  under air with a flow rate of  $20 \text{ ml min}^{-1}$ .

**Evolved gas analysis (EGA).** EGA experiments were performed using a Hiden CATLAB under the same conditions used in the TGA experiments.

**Copper surface area analysis.** This was determined by reactive frontal chromatography. Catalysts were crushed and sieved to a particle size of  $0.6\text{--}1 \text{ mm}$  and packed into a reactor tube. The catalyst was purged under helium for 2 minutes at  $70^\circ\text{C}$  before being heated under a  $5\% \text{ H}_2$  reduction gas to  $230^\circ\text{C}$  ( $8^\circ\text{C min}^{-1}$ ) for 3 hours. The catalyst was then cooled to  $68^\circ\text{C}$  under helium before the dilute  $2.5\% \text{ N}_2\text{O}$  reaction gas was added with a flow rate of  $65 \text{ ml min}^{-1}$ . The formation of  $\text{N}_2$  from the surface oxidation of the copper catalyst by  $\text{N}_2\text{O}$  was measured downstream using a thermal conductivity detector (TCD). Once the surface of the copper is fully oxidized, there is a breakthrough of  $\text{N}_2\text{O}$ , detected on the TCD. From this, the number of oxygen atoms that are chemisorbed on the copper surface can be determined. The number of exposed surface copper atoms and the copper surface area can then be derived. Quoted surface areas are calculated using discharged sample mass. Note that recent work<sup>37,38</sup> has shown that, if the catalyst is exposed to partial pressures of  $\text{H}_2$  exceeding  $0.05 \text{ bar}$ , then partial reduction of  $\text{ZnO}$  at the copper interface can occur. This will affect copper surface area results, owing to  $\text{N}_2\text{O}$  oxidizing both copper and  $\text{ZnO}$ . In these cases, alternative techniques, such as a  $\text{H}_2$  thermal conductivity detector, will give more accurate data with respect to copper surface area<sup>37,38</sup>.

**Copper surface area analysis after exposure to LTS conditions.** Copper surface area analysis of the fresh catalysts were carried out on a Quantachrome ChemBET 3000. Sample ( $100 \text{ mg}$ ) was packed into a stainless-steel U-tube and purged with high-purity helium for 5 minutes. Samples were reduced using  $10\% \text{ H}_2/\text{Ar}$  ( $30 \text{ ml min}^{-1}$ ) heated to  $140^\circ\text{C}$  at  $10^\circ\text{C min}^{-1}$ , before heating further to  $225^\circ\text{C}$  at  $1^\circ\text{C min}^{-1}$ . The resulting catalyst was held at this temperature for 20 minutes to ensure that complete reduction took place. Note that, under this partial pressure of  $\text{H}_2$ , it has been reported that  $\text{ZnO}$  species in contact with copper can partially reduce<sup>37,38</sup>. Residual  $\text{H}_2$  was flushed from the system by switching the gas line back over to helium ( $80 \text{ ml min}^{-1}$ ), while holding the sample at  $220^\circ\text{C}$  for another 10 minutes.

The temperature was then reduced to 65 °C for N<sub>2</sub>O pulsing (BOC AA Grade). A programme of 12 pulses of 113 µl N<sub>2</sub>O with a 5-minute stabilization time between pulses was carried out, followed by three pulses of N<sub>2</sub> for calibration. Unreacted N<sub>2</sub>O was trapped before reaching the detector using a molecular sieve 5A (pelleted, 1.6 mm, Sigma Aldrich) trap. The copper surface area was determined from the amount of N<sub>2</sub> emitted and the post-reaction analysis of catalyst mass, as follows:

$$\text{Copper surface area (m}^2\text{ g}^{-1}\text{)} = \text{N}_2 \text{ volume (ml)} \times N_A \times 2 / \text{Catalyst mass (g)} \times 24,000 \text{ (ml)} \times (1.47 \times 10^{19}) \text{ (atoms m}^{-2}\text{)}$$

where  $N_A$  is the Avogadro constant,  $6.022 \times 10^{23}$  (atoms). The key assumptions are that the amount of N<sub>2</sub> emitted amounts to half a monolayer's coverage of oxygen, and that the surface density of copper is  $1.47 \times 10^{19}$  (atoms m<sup>-2</sup>). The volume of N<sub>2</sub> produced was quantified using a TCD.

After copper surface area analysis, the samples were kept under N<sub>2</sub> and transferred to an LTS reactor. Ageing was carried out in a single fixed-bed reactor equipped with a bypass line. CO, N<sub>2</sub>, CO<sub>2</sub> and H<sub>2</sub> were introduced to the catalyst bed via mass-flow controllers (Bronkhorst). Water of high-performance liquid chromatography (HPLC) grade was passed through a liquid-flow controller (Bronkhorst) and then into a controlled evaporator mixer (Bronkhorst) that was heated to 140 °C. N<sub>2</sub> was fed through the vaporized water to give a dilute syngas mixture (H<sub>2</sub>O/CO/CO<sub>2</sub>/H<sub>2</sub>/N<sub>2</sub> = 25/1/4/13.75/56.25). This mixture was introduced at 220 °C after re-reduction of the catalyst. The gas flows were controlled to achieve an MHSV of 30,000 h<sup>-1</sup> kg<sup>-1</sup>. After ageing for 40 hours, the samples were transferred back to the Quantachrome ChemBET Chemisorption analyser, whereby, after re-reduction, the copper surface areas of the aged catalysts were measured, as described above.

**Scanning transmission electron microscopy (STEM).** Samples for STEM characterization were dry-dispersed onto holey carbon TEM grids. They were examined in an aberration-corrected JEOL ARM-200CF scanning transmission electron microscope operating at 200 kV in bright-field and dark-field STEM imaging modes. Reliable electron microscopy results could be obtained only from the set of zincian-georgeite- and zincian-malachite-derived materials, as these were found to be stable under the vacuum environment of the microscope, and largely unaffected by electron-beam irradiation. By way of contrast, the copper-only georgeite precursor materials were highly unstable under vacuum conditions (even without electron-beam irradiation), turning from a blue to dark-green colour, probably owing to the loss of occluded water. The corresponding zincian-georgeite materials showed no such colour transformation under vacuum.

**Environmental transmission electron microscopy (ETEM).** Samples for reduction during characterization by ETEM were dry-dispersed on heater chips (DensSolution trough hole) and then mounted on a DensSolution SH30 heating holder. The holder with sample was inserted into an FEI Titan 80-300 environmental transmission electron microscope operated at 300 kV (ref. 39). The reduction of the samples was performed *in situ* as follows, for both the georgeite and the malachite precursors: a flow of H<sub>2</sub> was let into the microscope, building up a pressure of 2 mbar. The sample was heated from room temperature to 150 °C using a heating ramp rate of 10 °C min<sup>-1</sup>. The final heating to 225 °C was done at 1 °C min<sup>-1</sup>. The oxidation state of copper was monitored by electron energy-loss spectroscopy (EELS) during the reduction treatment using a Gatan Tridiem 866 spectrometer attached to the microscope. After an extended *in situ* treatment of several hours at 225 °C, phase-contrast lattice imaging was performed at elevated

temperature in an H<sub>2</sub> atmosphere for both the georgeite and malachite samples, and recorded using a Gatan US 1000 charge-coupled-device camera.

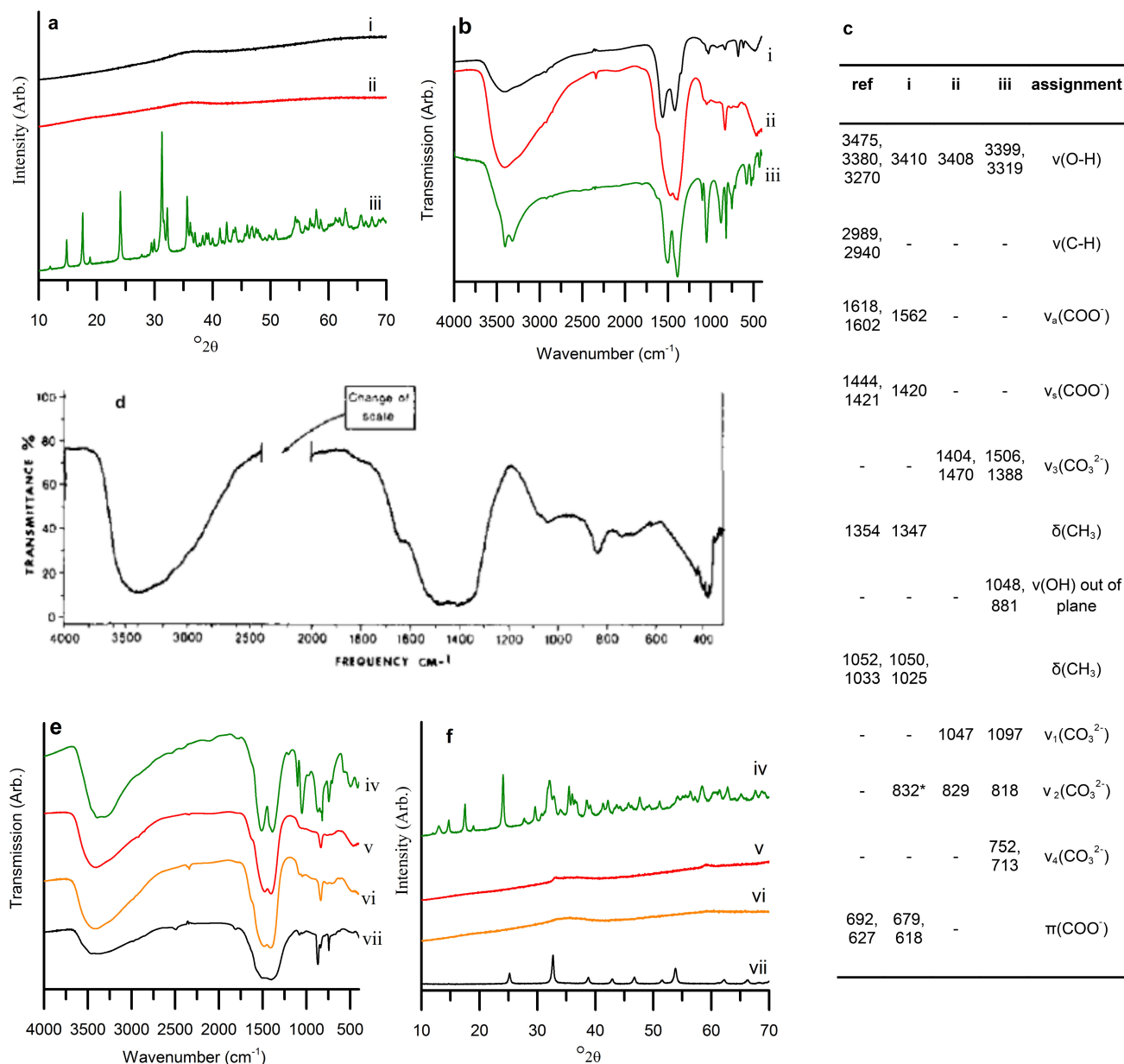
**X-ray photoelectron spectroscopy (XPS).** A Kratos Axis Ultra DLD system was used to collect XPS spectra, using a monochromatic Al K<sub>α</sub> X-ray source operating at 120 W. Data were collected with pass energies of 160 eV for survey spectra, and 40 eV for the high-resolution scans. The system was operated in the Hybrid mode, using a combination of magnetic immersion and electrostatic lenses and acquired over an area approximately 300 × 700 µm<sup>2</sup>. A magnetically confined charge-compensation system was used to minimize charging of the sample surface, and the resulting spectra were calibrated to the C(1 s) line at 284.8 eV; all spectra were taken with a 90° take-off angle. A base pressure of ~1 × 10<sup>-9</sup> Torr was maintained during collection of the spectra. Gas treatments were performed in a Kratos catalysis cell, which mimics the conditions of a normal reactor vessel, allowing the re-creation of reactor conditions and analysis of the chemical changes taking place on the catalyst surface. Briefly, the cell consists of a fused quartz reactor vessel contained within a stainless-steel vacuum chamber (base pressure ~10<sup>-8</sup> Torr after baking). Samples were heated at a controlled ramp rate of 2 °C min<sup>-1</sup> to a temperature of 225 °C using a eurotherm controller. The catalysts were exposed to an atmosphere of 2% H<sub>2</sub> in nitrogen with a flow rate of 30 ml min<sup>-1</sup>, controlled using MKS mass flow controllers during the heating ramp, during the 20-minute isotherm at 225 °C, and also while the catalyst was cooled to 25 °C. The samples were analysed before and after gas treatment without breaking the vacuum.

**Inductively coupled plasma mass spectrometry (ICP-MS) and carbon-hydrogen-nitrogen (CHN) analysis.** These analyses were provided as a commercial service by Warwick Analytical Services.

**Helium pycnometry.** This analysis was provided as a commercial service by MCA Services.

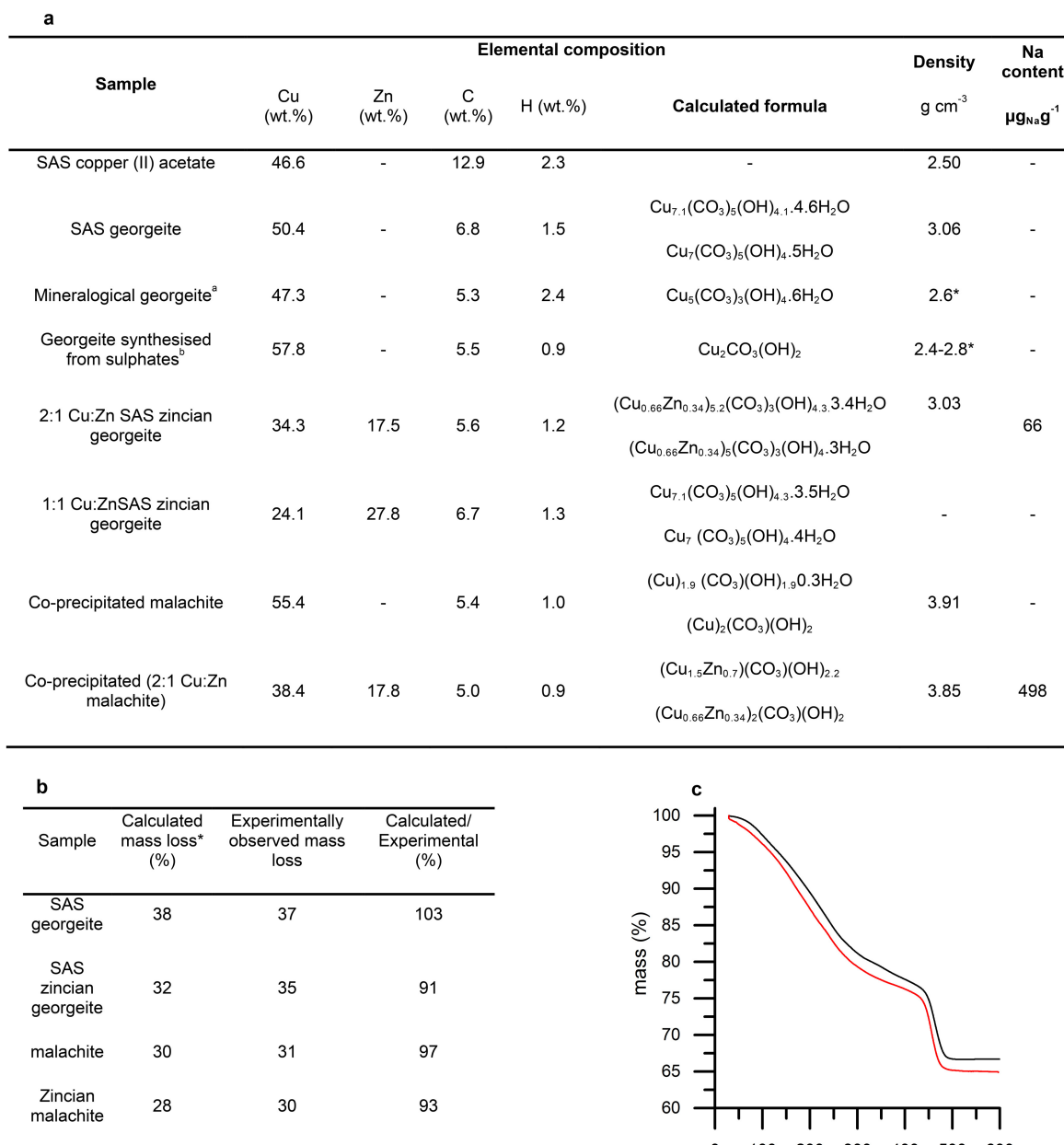
30. Coelho, A. A. TOPAS Academic: General Profile and Structure Analysis Software for Powder Diffraction Data (Bruker AXS, Karlsruhe, 2010).
31. Balzar, D. *et al.* Size-strain line-broadening analysis of the ceria round-robin sample. *J. Appl. Cryst.* **37**, 911–924 (2004).
32. Newville, M. IFEFFIT: interactive XAFS analysis and FEFF fitting. *J. Synchrotron Radiat.* **8**, 322–324 (2001).
33. Ravel, B. & Newville, M. ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT. *J. Synchrotron Radiat.* **12**, 537–541 (2005).
34. Chupas, P. J. *et al.* Rapid-acquisition pair distribution function (RA-PDF) analysis. *J. Appl. Cryst.* **36**, 1342–1347 (2003).
35. Soper, A. K. & Barney, E. R. Extracting the pair distribution function from white-beam X-ray total scattering data. *J. Appl. Cryst.* **44**, 714–726 (2011).
36. Keen, D. A. A comparison of various commonly used correlation functions for describing total scattering. *Angew. Chem. Int. Edn* **53**, 7043–7047 (2014).
37. Fichtl, M. B. *et al.* Counting of oxygen defects versus metal surface sites in methanol synthesis catalysts by different probe molecules. *J. Appl. Cryst.* **34**, 172–177 (2001).
38. Kuld, S., Conradsen, C., Moses, P. G., Chorkendorff, I. & Sehested, J. Quantification of zinc atoms in a surface alloy on copper in an industrial-type methanol synthesis catalyst. *Angew. Chem. Int. Edn* **53**, 5941–5945 (2014).
39. Hansen, T. W., Wagner, J. B. & Dunin-Borkowski, R. E. Aberration corrected and monochromated environmental transmission electron microscopy: challenges and prospects for materials science. *Mater. Sci. Technol.* **26**, 1338–1344 (2010).





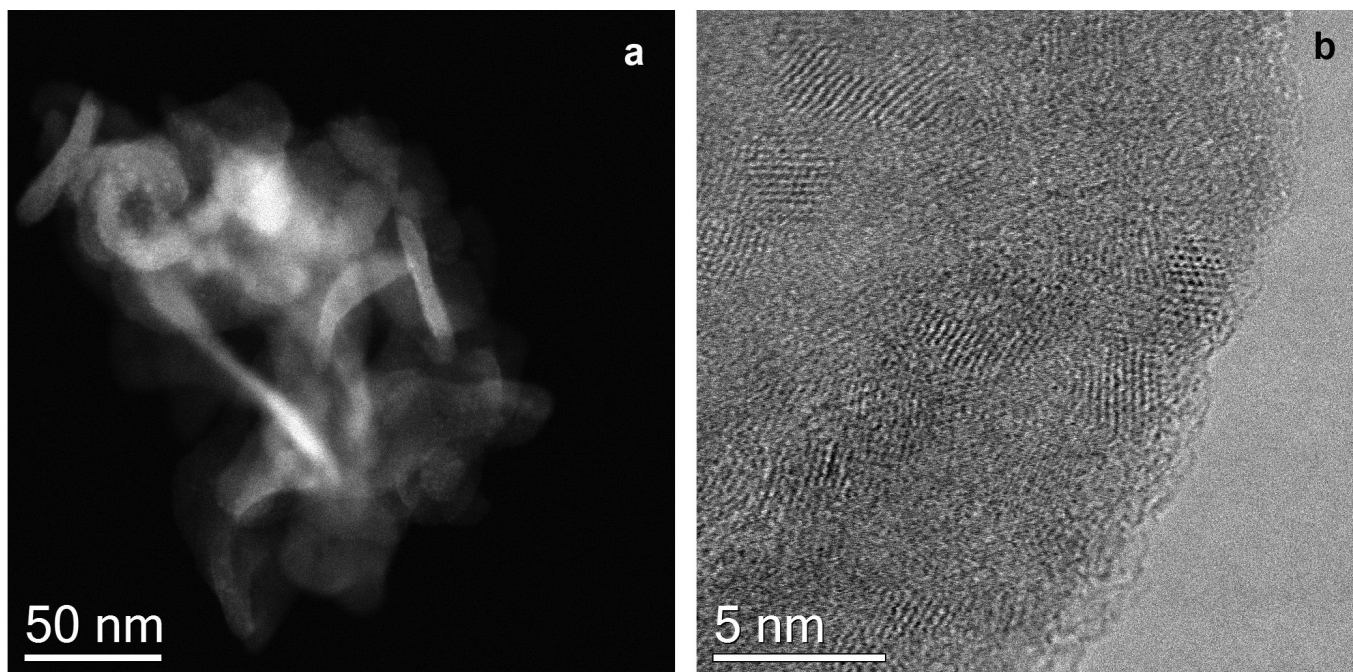
**Extended Data Figure 1 | FT-IR and XRD characterization of SAS copper and copper/zinc acetate precipitates.** Key: i, SAS-prepared copper acetate; ii, SAS-prepared georgeite; iii, malachite prepared by co-precipitation; iv, 2/1 copper/zinc malachite prepared by co-precipitation; v, SAS-prepared 2/1 copper/zinc georgeite; vi, SAS-prepared 1/1 copper/zinc georgeite; vii, SAS-prepared smithsonite ( $\text{ZnCO}_3$ ). **a**, XRD analysis of copper-only samples. **b**, FT-IR analysis of copper-only samples. **c**, FT-IR band assignment of copper-only samples, with (reference) designated as received copper(II) acetate monohydrate. Values given are for IR band

positions in wavenumbers ( $\text{cm}^{-1}$ ). \*The presence of this band shows that SAS precipitation with no additional water co-solvent produced some georgeite as well as amorphous copper acetate. We attribute the formation of georgeite to the small amount of water present from the monohydrated starting salt. **d**, IR spectrum of mineralogical georgeite, reproduced with the permission of the Mineralogical Society of Great Britain and Ireland from ref. 11. **e**, FT-IR spectra of copper/zinc samples. **f**, XRD of copper/zinc samples.



**Extended Data Figure 2 | Elemental composition of copper and copper/zinc samples, with supplementary TGA analysis.** **a**, Elemental composition of SAS-prepared georgeite and co-precipitated malachite samples. Elemental composition was determined by CHN analysis and ICP-MS. Densities were determined by helium pycnometry. <sup>a</sup>Values from ref. 11. <sup>b</sup>Values from ref. 16. \*Density determined by sink–float (SF) method. The helium pycnometry used in our present study provides

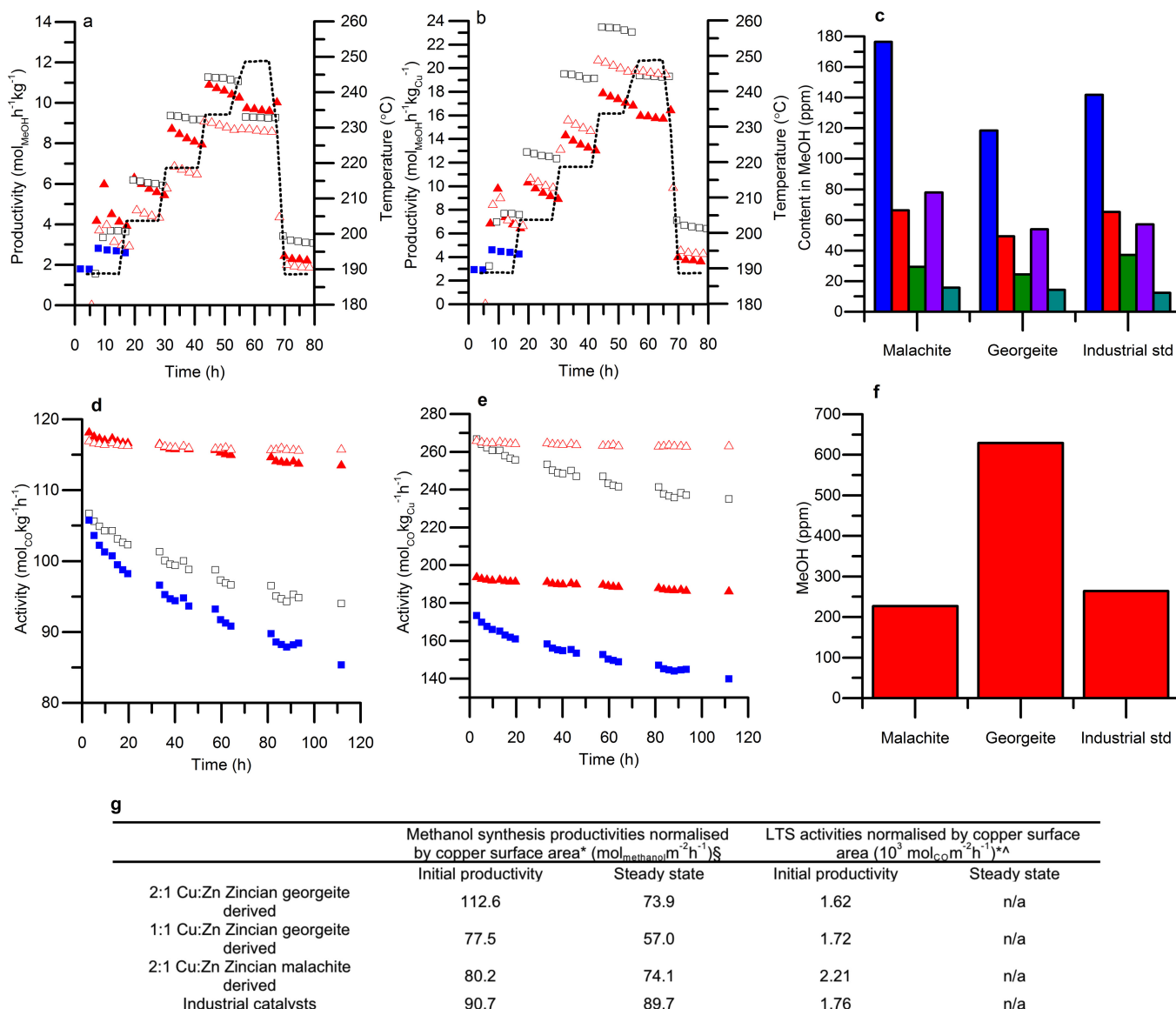
a skeletal density that negates buoyancy effects. **b**, Comparison of experimental and calculated mass losses for georgeite and malachite from TGA measurements. \*Calculated from the assumption of final products being 2/1 copper oxide/zinc oxide. The lower-than-expected mass losses for the zincian phases could be associated with the inclusion of copper ions in the zinc oxide lattice. **c**, TGA analysis of 2/1 copper/zinc georgeite (red line) and 1/1 copper/zinc georgeite (black line).



**Extended Data Figure 3 | Representative STEM micrographs of the zincian georgeite precursor.** Left, dark-field (DF) STEM micrograph. Right, bright-field (BF) micrograph. The general morphology of the zincian-georgeite precursor is shown in the DF-STEM micrograph. It typically consists of very characteristic, irregularly shaped agglomerates, about 100–200 nm in diameter, that are composed of ‘amorphous’ non-faceted particles about 40 nm wide. Closer inspection by BF-STEM shows that these non-faceted particles consist of an amorphous matrix phase in which are embedded largely disconnected, sub-2-nm crystallites of ordered material exhibiting clear lattice fringes. The amorphous matrix, probably containing the carbonate and hydroxyl species, is by far the

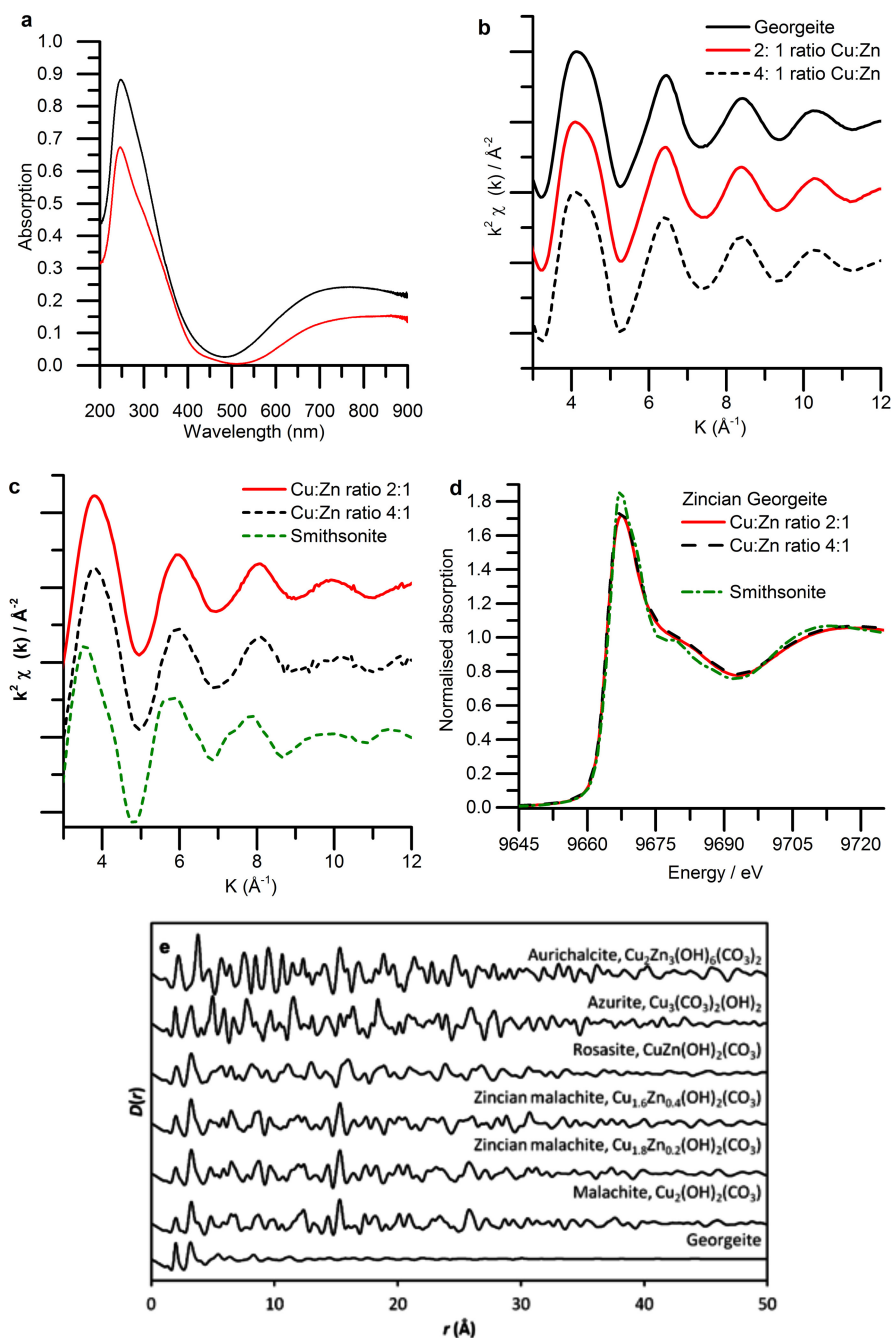
majority phase, while the nanocrystallites make up less than 10% of the material by volume. This observation is consistent with our other characterization data, as signals from the matrix phase would dominate the XAFS analysis, whereas the nanocrystallites are too small in dimension to be detected by XRD. Analysis of fringe spacings and interplanar angles of individual nanocrystallites from such images suggests that some of the grains could be CuO (tenorite, where the copper is fourfold coordinated by oxygen), while others fit better to Cu<sub>2</sub>O (cuprite, where the copper has a coordination number of 2). No convincing matches of the lattice fringes to either ZnO or ZnCO<sub>3</sub> could be found.





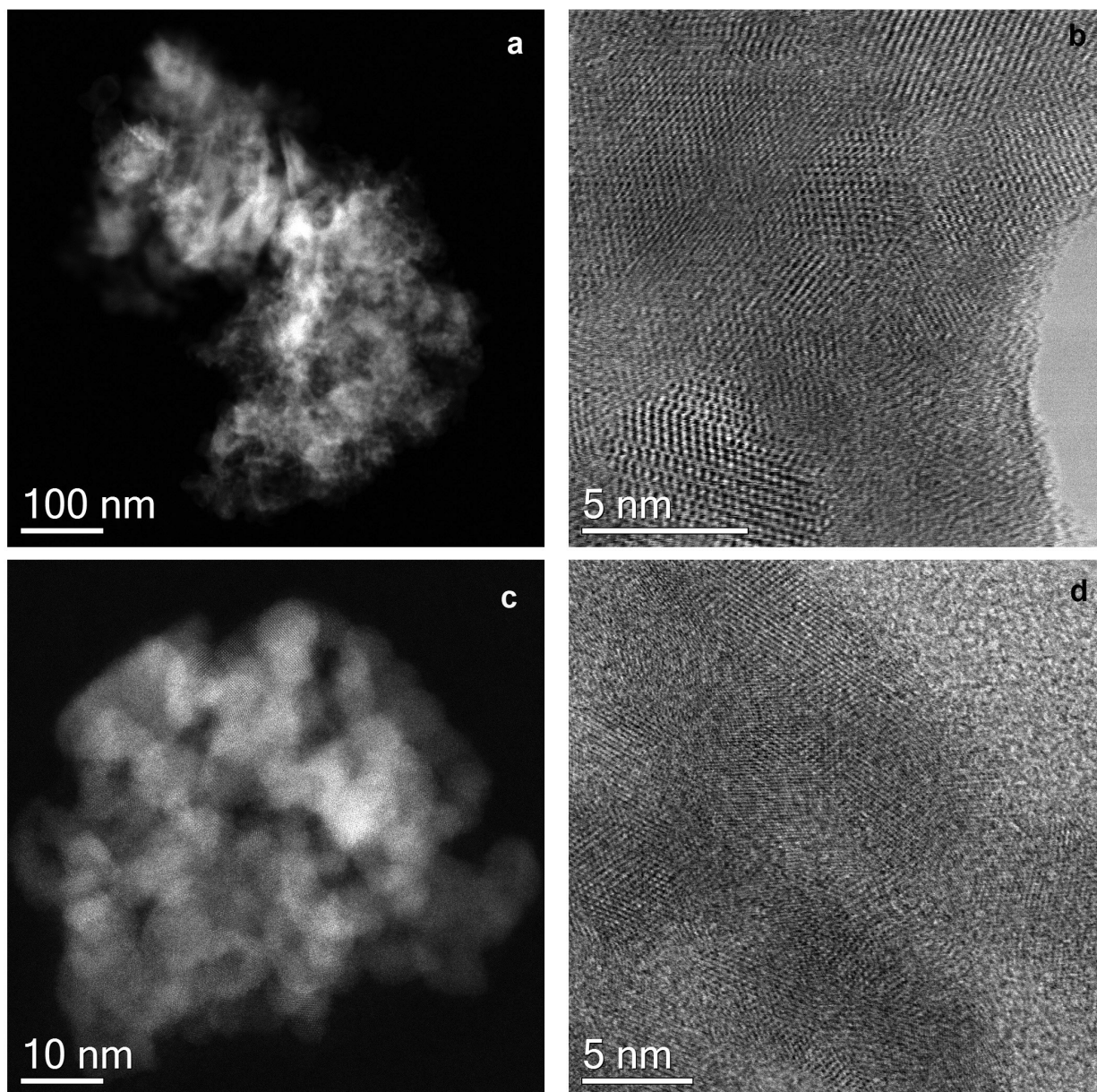
**Extended Data Figure 4 | Further catalyst testing in methanol-synthesis and LTS reactions.** Filled red triangles, 2/1 Cu/Zn georgeite; open triangles, 1/1 Cu/Zn georgeite; open squares, industrial standard; filled blue squares, 2/1 Cu/Zn malachite (this test was terminated after 190 °C because of poor activity). **a, b**, Methanol-synthesis data, normalized to total catalyst mass (**a**) and copper mass (**b**), at 190–250 °C. The dashed line shows the representative reactor-bed temperature. Reaction conditions were as follows: pressure 25 bar; gas composition  $\text{CO}/\text{CO}_2/\text{H}_2/\text{N}_2 = 6/9.2/67/17.8$ ;  $\text{MHSV} = 7,2001 \text{ kg}^{-1} \text{ h}^{-1}$ . **c**, Concentration of byproducts that were collected in the condensate pot after the methanol-synthesis reaction, as determined by gas-chromatographic (GC) analysis. Byproducts are: ethanol (blue), propanol (red), butanol (green), *iso*-butanol (purple), and methyl *iso*-butyl ketone (turquoise). **d, e**, LTS-reaction data, normalized to total catalyst mass (**d**) and copper mass (**e**). Reaction conditions were as follows: temperature 220 °C; pressure

27.5 bar; gas composition  $\text{H}_2\text{O}/\text{CO}/\text{CO}_2/\text{H}_2/\text{N}_2 = 50/2/8/27.35/12.5$ ;  $\text{MHSV} = 75,0001 \text{ kg}^{-1} \text{ h}^{-1}$ . **f**, Concentration of methanol collected in the condensate pot after the LTS reaction, as determined by GC analysis. **g**, Methanol-synthesis productivities and LTS activities, normalized by copper surface area. \*Copper surface area analysis determined by  $\text{N}_2\text{O}$  reactive frontal chromatography before testing. <sup>§</sup>Methanol-synthesis data acquired at 190 °C, with steady state being at 18 hours' time-on-line. <sup>^</sup>LTS data acquired at 220 °C, with steady state being at 40 hours time-on-line. Note that LTS simulation testing showed that copper surface area dropped markedly after 40 hours (to  $17 \text{ m}^2 \text{ g}^{-1}$  and  $19 \text{ m}^2 \text{ g}^{-1}$  for zincian-georgeite-derived and zincian-malachite-derived catalysts, respectively). The inverse correlation between copper surface area and initial productivity also suggests that loss of surface area occurs rapidly during the reaction and that the initial rate data are therefore likely to be inaccurate.



**Extended Data Figure 5 | Spectroscopic analysis of the addition of zinc to georgeite.** **a**, Diffuse reflectance UV-vis spectra of zincian malachite (red) and zincian georgeite (black). **b**, Copper K-edge EXAFS ( $\chi$ ) comparison of zincian georgeite (with a 4/1 or 2/1 copper/zinc ratio) with georgeite. **c**, Zinc K-edge XANES comparison of zincian georgeite

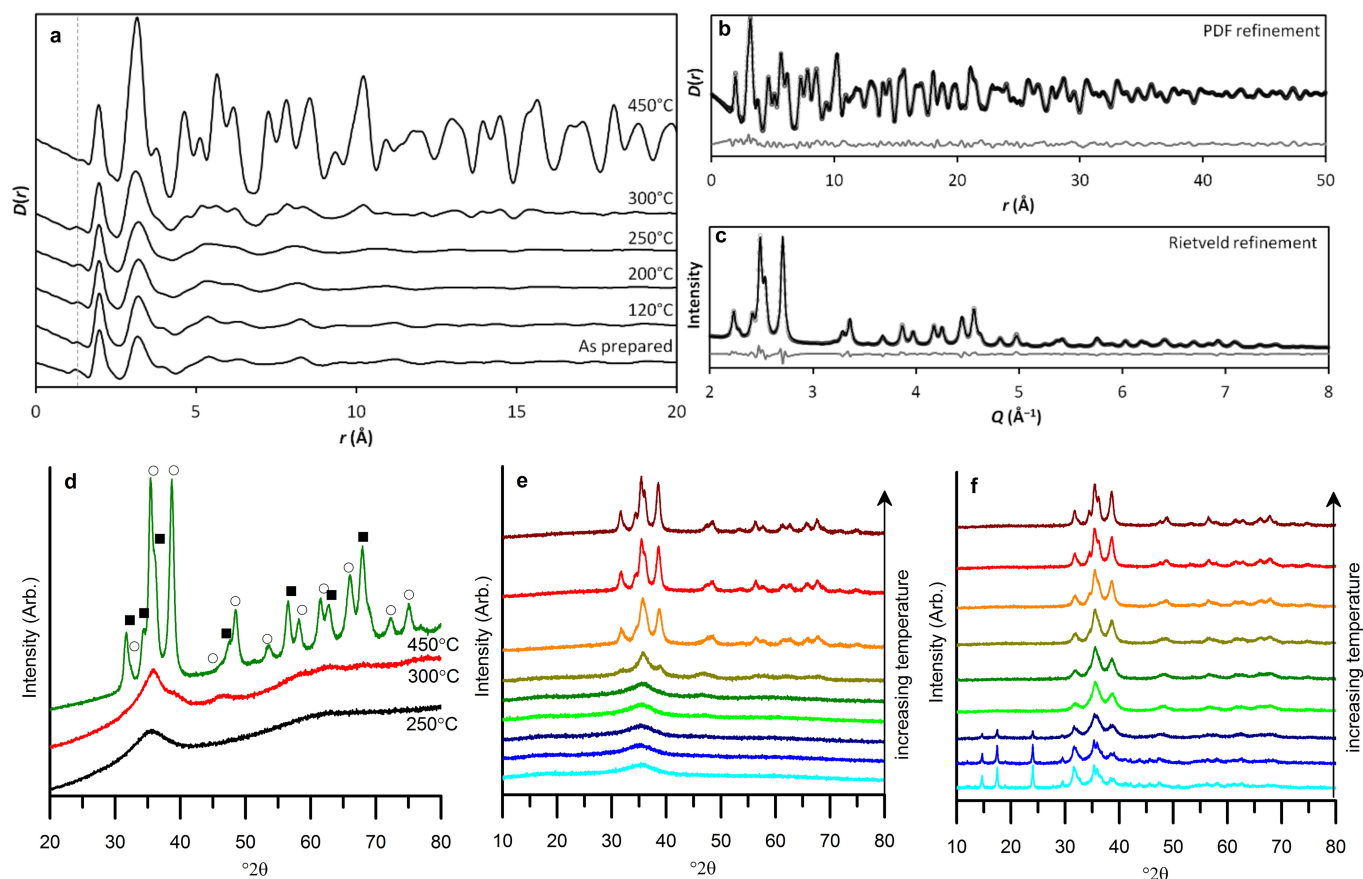
with SAS-prepared smithsonite. **d**, Zinc K-edge EXAFS ( $\chi$ ) comparison of zincian georgeite with SAS-prepared smithsonite. **e**, Comparison of observed georgeite PDF data with simulated data for crystalline hydroxycarbonate minerals with similar compositions to that of georgeite—namely aurichalcite, azurite, rosasite, zincian malachite and malachite.



**Extended Data Figure 6 | Representative DF-STEM and BF-STEM micrographs of zincian georgeite and zincian malachite, calcined at 300 °C.** **a**, DF-STEM of calcined zincian georgeite. Higher-magnification imaging reveals that, after much of the carbonate and hydroxyl content is lost by calcination, most of the disordered matrix material in the precursor has crystallized, and only a small amount of amorphous material remains. The crystallized material is entirely in a nanocrystalline form, with a mean

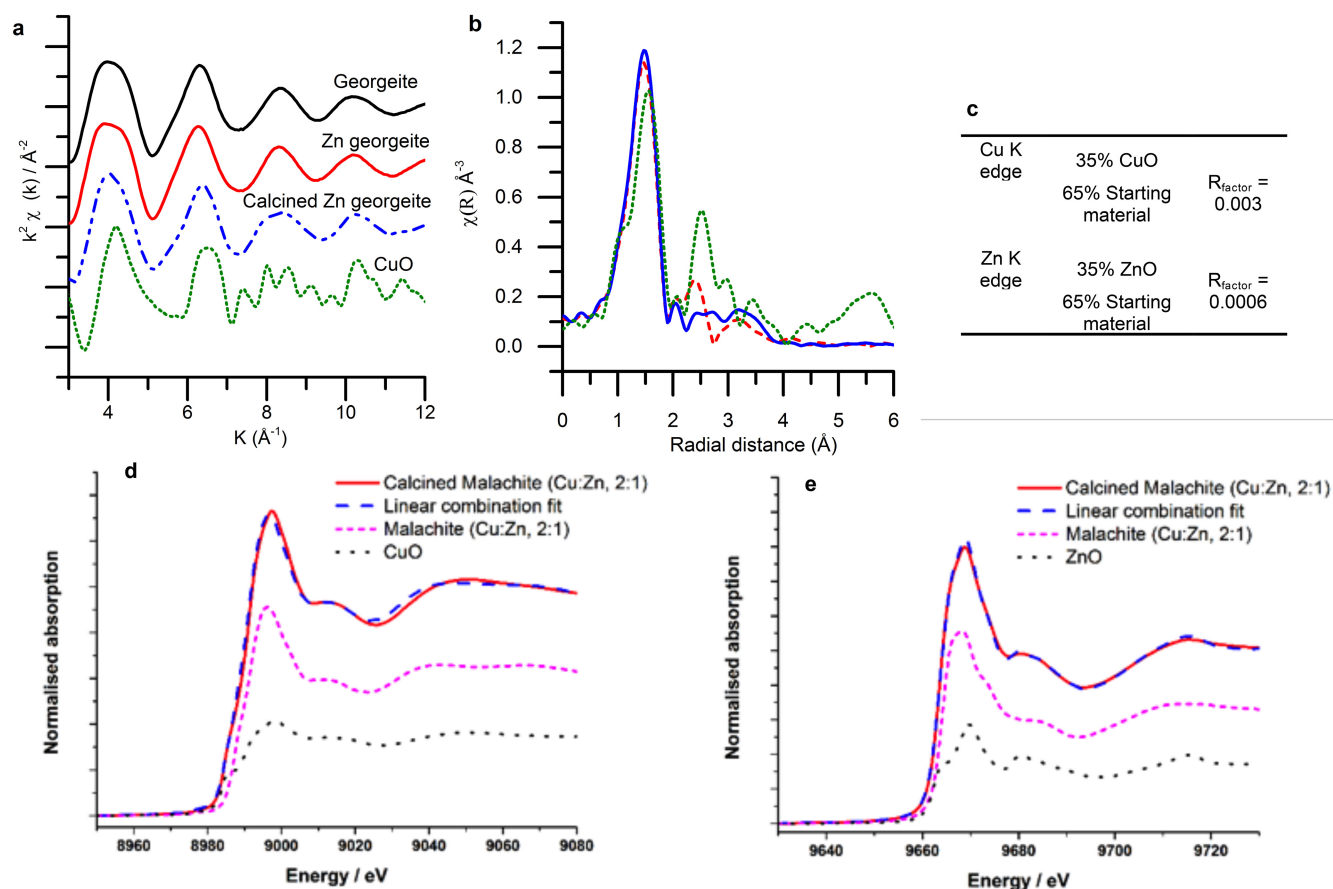
grain diameter of 3–4 nm, which is just below the detection limit for XRD. Analysis of the fringe spacings and interplanar angles from individual grains suggests that the material is now mainly an intimate mixture of zinc and copper oxides; the small amount of disordered material corresponds to the residual occluded carbonate material, as detected by TGA/EGA analysis of this material. **b**, BF-STEM of calcined zincian georgeite. **c**, DF-STEM of calcined zincian malachite. **d**, BF-STEM of calcined zincian malachite.





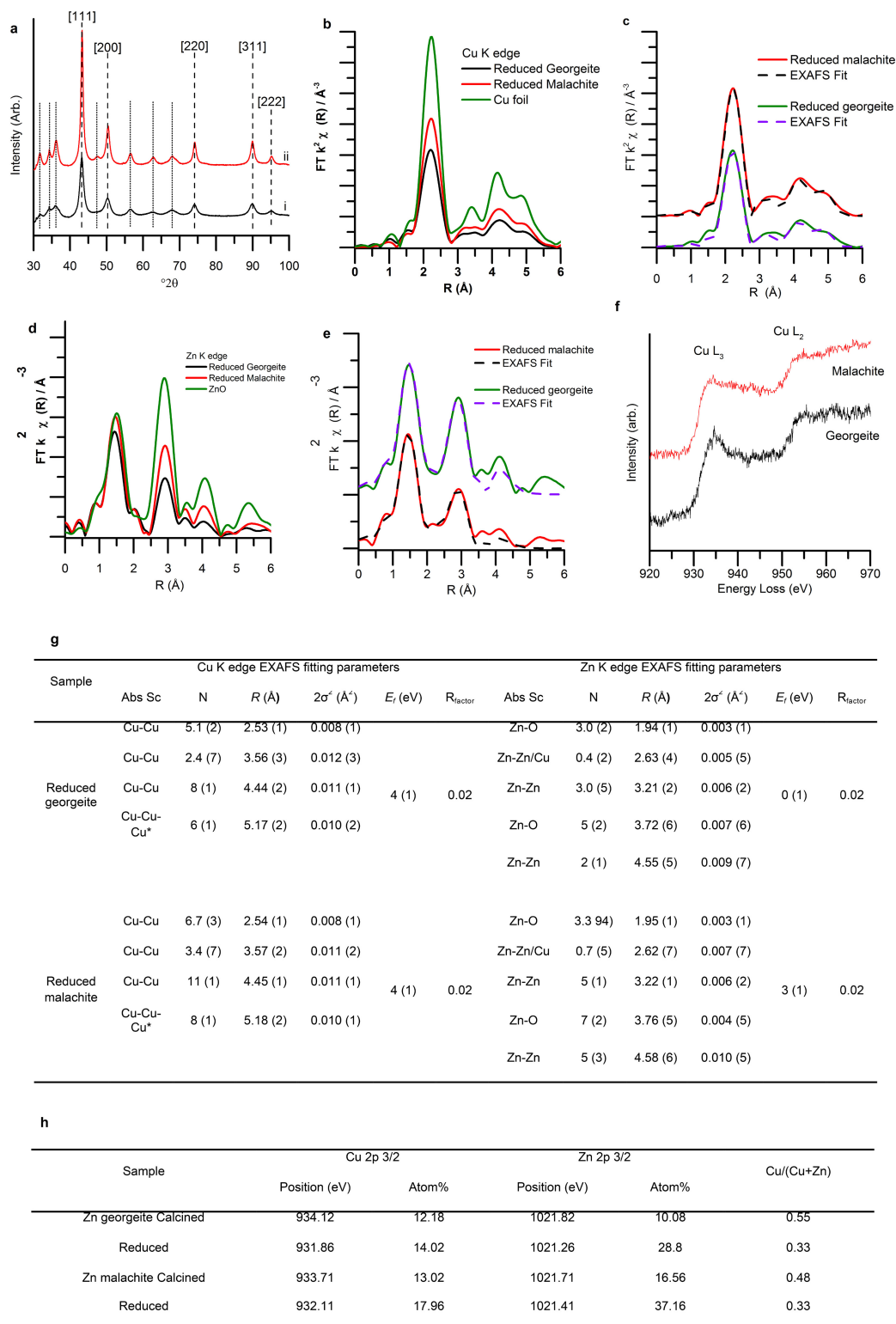
**Extended Data Figure 7 | X-ray diffraction analysis of calcined zincian georgeite and zincian malachite.** **a**, PDF of zincian georgeite, as prepared and after calcination at 120 °C, 200 °C, 250 °C, 300 °C and 450 °C. There is little change in the observed PDF up to 250 °C, other than a slight peak broadening, which can be attributed to a reduction in short-range order. The dashed line shows the position of the C–O peak, which is retained until temperatures higher than 300 °C. **b**, PDF and **c**, Rietveld fits of zincian georgeite after calcination at 450 °C. The measured data are shown

as open circles; the fit is a solid black line; the difference is a grey line. Both techniques determine the product to be a mixture of copper oxide and zinc oxide (weight ratio of 68/32 by PDF; 67.7(4)/32.3(4) by Rietveld). **d**, *Ex situ* XRD patterns following calcination of zincian georgeite for 2 hours at 250 °C, 300 °C and 450 °C. Open circles, copper oxide; filled squares, zinc oxide. **e**, **f**, *In situ* XRD analysis of zincian georgeite (**d**) and zincian malachite (**e**) during calcination between 300 °C and 500 °C under an atmosphere of static air, with XRD scans every 25 °C.



**Extended Data Figure 8 | Copper K-edge XAFS analysis of zincian georgeite and zincian malachite calcined at 300 °C.** **a**, EXAFS ( $\chi$ ) comparison of copper oxide, calcined zincian georgeite, zincian georgeite and georgeite. **b**, EXAFS ( $R$ ) comparison of copper oxide (green),

calcined zincian georgeite (black) and zincian georgeite (blue). **c–e**, Linear combination fit of copper oxide and zinc oxide with zincian malachite calcined at 300 °C.



**Extended Data Figure 9 | *In situ* characterization of final-state, reduced copper/zinc-oxide catalysts derived from zincian georgeite and zincian malachite, calcined at 300 °C.** **a**, XRD of zincian-georgeite-derived catalyst after *in situ* hydrogen reduction (2% H<sub>2</sub>/N<sub>2</sub> at 225 °C for 1 hour). Fine dotted lines indicate zinc-oxide reflections; dashed lines indicate metallic copper. **b**, Copper K-edge EXAFS Fourier transform of final reduced catalysts derived from zincian georgeite and zincian malachite. **c**, Copper K-edge EXAFS fit of reduced catalysts. **d**, Zinc K-edge EXAFS Fourier transform of final reduced catalysts derived from zincian georgeite and zincian malachite. **e**, Zinc K-edge EXAFS fit of reduced catalysts. **f**, Electron energy-loss spectra (EELS) of georgeite and malachite, respectively, during reduction in the ETEM experiment. The EELS data

were acquired in 2 mbar H<sub>2</sub> at 225 °C, and show the fine structure of the Cu L<sub>2,3</sub> ionization edges, which are characteristic of metallic Cu<sup>0</sup>. **g**, EXAFS fitting data for copper and zinc K-edge data. Fitting parameters for K-edge copper: amplitude-reduction factor ( $S_0^2$ ) = 0.91, as deduced from copper-foil standard; fit range  $3 < k < 11.2$ ,  $1 < R < 5.5$  (with  $k$  denoting the fitting window from the  $\chi$  data, and  $R$ , the path length, denoting the fitting from the Fourier transform of the  $\chi$  data); number of independent points = 23; \*denotes multiple scattering path. Fitting parameters for K-edge Zn:  $s_0^2$  = 0.90 as deduced from a ZnO<sub>2</sub> standard; fit range  $3.3 < k < 9.5$ ,  $1 < R < 4.8$ ; number of independent points = 16; \*denotes multiple scattering path. **h**, XPS analysis of calcined and reduced catalysts derived from zincian georgeite and zincian malachite.



Extended Data Table 1 | EXAFS fitting parameters for georgeite and malachite

Abs Sc	<i>N</i>	<i>R</i> (Å)	$2\sigma^2$ (Å <sup>2</sup> )	<i>E<sub>r</sub></i> (eV)	<i>R<sub>factor</sub></i>
Malachite					
Cu–O	4 (fixed)	1.92 (2)	0.007 (2)	-1(1)	0.014
Cu–O	2 (fixed)	2.45 (7)	0.02 (1)		
Cu–C	3 (fixed)	3.01 (4)	0.002 (6)		
Cu–Cu	2 (fixed)	3.12 (4)	0.006 (5)		
Cu–Cu	3 (fixed)	3.31 (3)	0.007 (3)		
Georgeite					
Cu–O	4 (fixed)	1.94 (1)	0.004 (1)	2(1)	0.007
Cu–O	2 (fixed)	2.45 (fixed)	0.03 (fixed)		

Fitting parameters for malachite:  $S_0^2 = 0.72$  as deduced from a copper-oxide standard; fit range  $3 < k < 12$ ,  $1 < R < 3.5$ ; number of independent points = 14. Fitting parameters for georgeite:  $s_0^2 = 0.72$  as deduced from a copper-oxide standard; fit range  $3 < k < 12$ ,  $1 < R < 3.5$ ; number of independent points = 14. Abs Sc; scattering path, *N*; coordination number (first shell); *R*, path length,  $\sigma^2$ ; Debye–Waller factor, *E<sub>i</sub>*; change in edge energy, *R<sub>factor</sub>*; goodness of fit.

# Electrostatic catalysis of a Diels–Alder reaction

Albert C. Aragonès<sup>1,2,3</sup>, Naomi L. Haworth<sup>4</sup>, Nadim Darwish<sup>1,2</sup>, Simone Ciampi<sup>5</sup>, Nathaniel J. Bloomfield<sup>4</sup>, Gordon G. Wallace<sup>5</sup>, Ismael Díez-Pérez<sup>1,2,3</sup> & Michelle L. Coote<sup>4</sup>

**It is often thought that the ability to control reaction rates with an applied electrical potential gradient is unique to redox systems. However, recent theoretical studies suggest that oriented electric fields could affect the outcomes of a range of chemical reactions, regardless of whether a redox system is involved<sup>1–4</sup>. This possibility arises because many formally covalent species can be stabilized via minor charge-separated resonance contributors. When an applied electric field is aligned in such a way as to electrostatically stabilize one of these minor forms, the degree of resonance increases, resulting in the overall stabilization of the molecule or transition state. This means that it should be possible to manipulate the kinetics and thermodynamics of non-redox processes using an external electric field, as long as the orientation of the approaching reactants with respect to the field stimulus can be controlled. Here, we provide experimental evidence that the formation of carbon–carbon bonds is accelerated by an electric field. We have designed a surface model system to probe the Diels–Alder reaction, and coupled it with a scanning tunnelling microscopy break-junction approach<sup>5–7</sup>. This technique, performed at the single-molecule level, is perfectly suited to deliver an electric-field stimulus across approaching reactants. We find a fivefold increase in the frequency of formation of single-molecule junctions, resulting from the reaction that occurs when the electric field is present and aligned so as to favour electron flow from the dienophile to the diene. Our results are qualitatively consistent with those predicted by quantum-chemical calculations in a theoretical model of this system, and herald a new approach to chemical catalysis.**

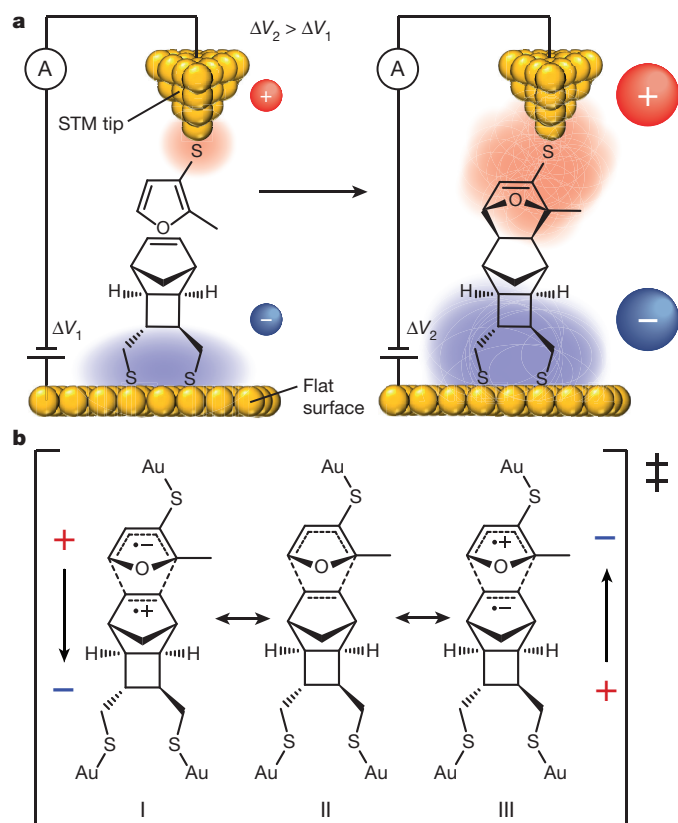
All chemical reactions can be viewed as the movement of electrons and/or nuclei; as such, one might expect that their kinetics and thermodynamics could be influenced by external electric fields. Even for non-redox reactions, theoretical studies predict that electrostatic effects should in principle influence the stabilities of chemical species by stabilizing or destabilizing charge-separated resonance contributors<sup>1–4</sup>. For example, Shaik and co-workers have argued that a covalent bond of the form X–Y can be thought of as comprising several possible resonance contributors,  $[XY \leftrightarrow X^+Y^- \leftrightarrow X^-Y^+]$ <sup>8,9</sup>. In the absence of an electric field, the extent to which either of the charge-separated structures will contribute to the resonance stabilization of the bond will depend on the relative electronegativities of X and Y. Indeed, Shaik and colleagues were able to use this concept to explain energy trends in the group 14 MH<sub>3</sub>–Cl bonds<sup>10</sup>, among other examples. In this context, the presence of an appropriately oriented external electric field has the potential to further stabilize or destabilize these charge-transfer contributors, and thereby influence bond energy. Moreover, the participation of minor charge-separated resonance structures is not limited to covalent-bond energies; it has been invoked to explain trends in kinetics and thermodynamics for a wide range of chemical reactions, including radical addition<sup>11</sup> and transfer<sup>12</sup>. Thus, in principle, the scope of electrostatic catalysis—manipulating chemical reactions with electric fields—should be broad.

Electrostatic catalysis is the least developed form of catalysis in synthetic chemistry (even though it is widely harnessed by enzymes<sup>13–16</sup>). This is because electrostatic effects are strongly directional and are effectively quenched in polar media. Enzymes overcome these problems by using a low-polarity active site, in which the substrate binds in a precisely oriented manner; one or more charged residues within this site can then create an oriented local electric field that can catalyse the reaction. In synthetic chemistry, one can mimic this process to some extent by using charged functional groups on the substrate or catalyst; however, balancing the need for low solvent polarity with the limited solubility of charged residues in non-polar solvents leads to compromises that weaken the catalytic effect. For example, aminoxyl radicals (R<sub>1</sub>R<sub>2</sub>NO•) are stabilized via resonance with R<sub>1</sub>R<sub>2</sub>N<sup>+</sup>O<sup>–</sup>. Thus, when a (remote) negatively charged functional group is placed on the left-hand side of the N–O• bond, the electrostatic stabilization of this minor contributor leads to further stabilization of the species<sup>2,3</sup>. This stabilization, which has been verified experimentally, promotes dissociation of the R<sub>1</sub>R<sub>2</sub>NO–H and R<sub>1</sub>R<sub>2</sub>NO–R bonds by as much as 20 kJ mol<sup>–1</sup> in the gas phase<sup>2</sup>; however, the effect, while still of practical significance as a ‘pH switch’ of radical stability, is effectively halved in energetic terms in low-polarity solvents such as dichloromethane<sup>17</sup>, and essentially quenched in polar solvents<sup>18</sup>.

If one could use external electric fields instead of charged chemical species as the ‘catalyst’, one could manipulate a much broader range of reactions, conveniently altering both reactivity and selectivity in a tunable manner that can be predicted by theory. However, to probe this concept experimentally, one must develop a method of controlling the orientation of the external electric field (EEF) with respect to the reaction centre. Previously, EEFs have been used to guide the selectivity of isomerization reactions in which polar intermediates or transition states are involved<sup>19,20</sup>; but controlling the orientation of the EEF as two molecules collide in bimolecular reactions adds another dimension to the problem. Here we show that this can be achieved by combining surface chemistry procedures with state-of-the-art single-molecule techniques that are based on scanning tunnelling microscopy (STM). STM-based single-molecule electrical measurements can reveal information on chemical coupling averaged over thousands of collisions. This gives us the ability to control the dynamics of the approaching reactants and deliver the field stimulus upon collision. Using this approach, we show that a simple, textbook bimolecular carbon–carbon bond-forming reaction, the Diels–Alder reaction—involving reagents of ostensibly negligible polarity—can be accelerated by an oriented EEF.

Diels–Alder reactions—which involve a conjugated diene and a substituted alkene (the ‘dienophile’)—constitute a major family of chemical processes that are used in the preparation of fine chemicals<sup>21</sup>. Our choice of these reactions was inspired by theoretical predictions by Shaik and colleagues<sup>1</sup>, who suggested that the barrier heights for certain Diels–Alder reactions can be lowered substantially when an electric field is oriented appropriately. Here we use a

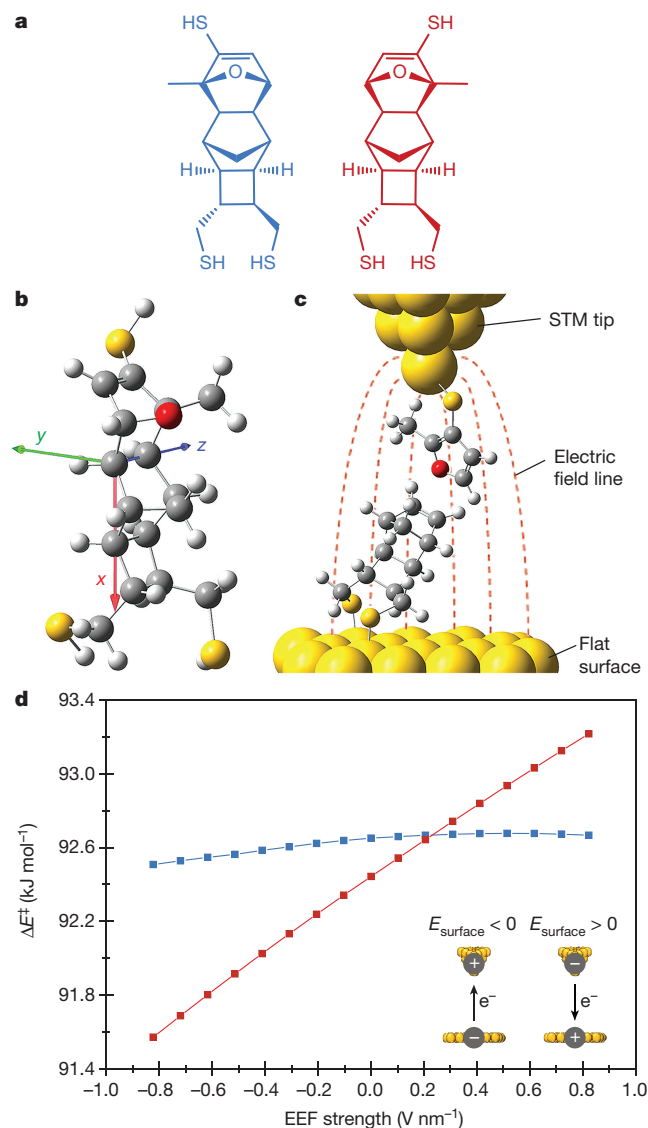
<sup>1</sup>Departament de Química-Física, Universitat de Barcelona, Diagonal 645, Barcelona 08028, Catalonia, Spain. <sup>2</sup>Institut de Bioenginyeria de Catalunya (IBEC), Baldri Reixac 15-21, Barcelona 08028, Catalonia, Spain. <sup>3</sup>Centro Investigación Biomédica en Red (CIBER-BBN), Campus Río Ebro-Edificio I+D, Poeta Mariano Esquillor s/n, Zaragoza 50018, Spain. <sup>4</sup>ARC Centre of Excellence for Electromaterials Science, Research School of Chemistry, Australian National University, Canberra, Australian Capital Territory 2601, Australia. <sup>5</sup>ARC Centre of Excellence for Electromaterials Science, Intelligent Polymer Research Institute, University of Wollongong, Wollongong, New South Wales 2500, Australia.



**Figure 1 | Electrostatic catalysis of a Diels–Alder reaction.** **a**, We studied the effects of an external electric field on the reaction rate by using single-molecule STM-BJ conductance measurements, which provide the oriented electric-field stimulus and also count reaction events. The diene (a furan) is attached to the STM tip via a thiol group ('S'); the dienophile (a norbornylogous bridge) is attached in a known orientation<sup>26</sup> to a flat gold surface via two thiols. Four structurally distinct products may be formed, each having two diastereoisomers; the kinetically favoured product is shown here.  $\Delta V$  is the voltage difference between the tip and surface electrodes. **b**, Possible resonance structures of the transition state. When an electric field is present, minor contributors I or III may be stabilized enough to undergo resonance with II, lowering the reaction barrier. The vertical arrows show the field direction most likely to stabilize I or III, with I expected to experience greater stabilization at a given field magnitude.

surface-tethered furan derivative as the diene, and a norbornylogous bridge with a terminal double bond as a non-polar dienophile ( $(\pm)$ -NB, tricyclo[4.2.1.0<sup>2,5</sup>]non-7-ene-3,4-dimethanethiol; only the (1*R*),2(*R*),3(*R*),4(*R*),5(*S*),6(*S*)-enantiomeric form is shown in Fig. 1a). Norbornylogous bridges are conformationally rigid molecules, and have been extensively used as electrical conduits for probing how geometrical and structural factors influence chemical and electrochemical phenomena<sup>22–24</sup>. The NB in Fig. 1a is a short, rigid, non-polar dienophile with two CH<sub>2</sub>SH groups (feet) in *trans*-stereochemistry; the presence of these feet allows unambiguous orientation of the distal double bond (dienophile) when assembled on flat gold surfaces<sup>23,25</sup>. The rigidity of NB helps us to position and align the dienophile with respect to the EEF when the diene part of the system is brought nearby. The molecular length of the dienophile is kept to a minimum (five sigma bonds) in order to maintain the Diels–Alder product within the conductance limit of our system.

To ascertain whether this Diels–Alder reaction is sensitive to the presence of an oriented EEF, we used quantum chemistry to study the field effect on the reaction barrier. This reaction has four structurally distinct Diels–Alder products; results for the reaction with the lowest barrier (*exo-syn*) are shown in Fig. 2 (see Supplementary Information for full results). There are two diastereoisomers for each of the four products, with the substituents of the furan being located either on the

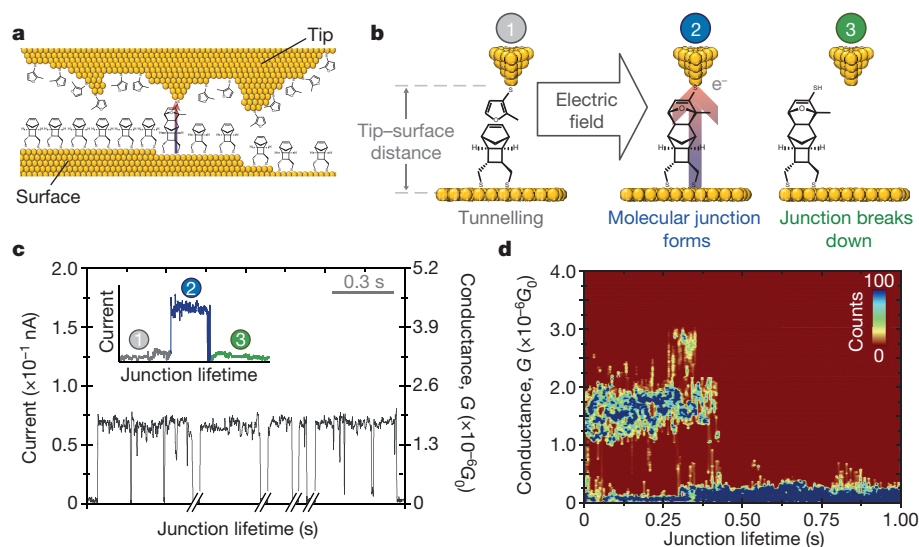


**Figure 2 | Computational modelling of the Diels–Alder reaction.** **a**, The two diastereoisomers of the *exo-syn* product of this reaction. These were the kinetically favoured products; the six other possible products had much higher reaction barriers over the experimental range of field strengths. In the blue diastereoisomer, the substituents of the furan are located on the left of the molecule; in the red diastereoisomer, these substituents are located on the right. **b**, The coordinate axes used to orient the field with respect to the molecule. The z axis lies along the double bond of the dienophile, while the x axis is directed along the NB backbone. **c**, The scenario being modelled, showing the NB bridge sitting in the experimentally determined orientation with respect to the surface of the STM plate and to the electric field lines, which are passing through the reaction centre at an oblique angle to the NB double bond. **d**, The predicted effects of the strength and direction of the external electric field (EEF) on the reaction-barrier height ( $\Delta E^\ddagger$ ) for formation of the two *exo-syn* diastereoisomers in **a** (see Supplementary Information). The formation of the blue structure is quite insensitive to the EEF over the experimental range of field strengths, while the formation of the red structure shows strong field sensitivity. The diagrams in the bottom right corner show the possible directions of the electric field (where  $E_{\text{surface}}$  denotes the voltage applied between the surface and the tip).

left of the molecule (see the blue diastereoisomer in Fig. 2a) or on the right (the red diastereoisomer).

The positioning of these substituents leads to different interactions with the CH<sub>2</sub>SH groups at the opposite end of the NB, resulting in slightly different energies when no EEF is present and very different responses to the applied field. Experimentally, the NB is known to sit at





**Figure 3 | Blinking experiments.** **a**, Diagram of the STM tip and surface during a blinking experiment. The STM tip was modified by furan molecules and the surface with NB molecules, using self-assembled-monolayer procedures (see Supplementary Information). **b**, The stages encountered during a blinking event. **c**, A real-time data capture of blinking events. The time breaks in the x axis are about 2 min. The inset shows the STM current response before (1), during (2) and after (3) the formation of a single blink (junction). **d**, Two-dimensional maps overlaying hundreds of blinks. Counts have been normalized to a colour scale, with 100 counts representing the maximum, and 0 representing the minimum. The surface bias was  $-0.5$  V in **c**, **d**.  $G_0 = 2e^2/h = 77.5 \mu\text{S}$ , with  $h$  the Planck constant and  $e$  the elementary charge.

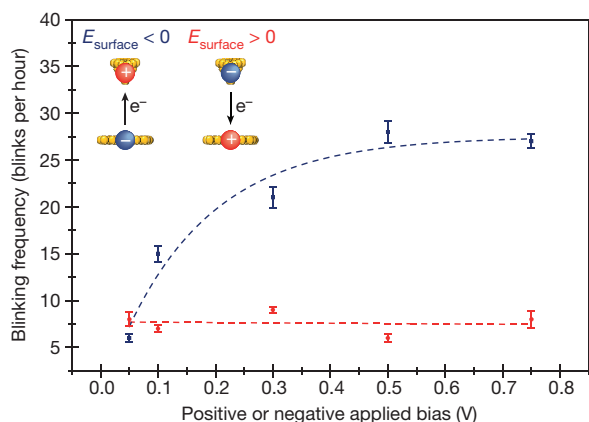
an angle to the gold surface<sup>26</sup>, tilted by  $30^\circ$  along the  $y$  axis and  $25^\circ$  along the  $z$  axis (Fig. 2c). For the *exo-syn* product, the  $30^\circ$   $y$ -axis tilt means that the field lines are oriented roughly along the average vector of the forming bonds. However, the  $z$ -axis tilt means that the furan SH group sits above the forming bonds for the blue diastereoisomer (Fig. 2c), but to the side of the molecule for the red diastereoisomer. As a result, the blue structure is predicted to be quite insensitive to the EEF over the experimental range of field strengths, while the red structure should show strong field sensitivity (see Supplementary Information, section 3). Specifically, it is predicted that, for negative bias, the barrier to formation of the red isomer will decrease with increasing field strength; for positive bias, it is predicted to increase (Fig. 2d). This trend occurs because the negatively biased EEF can stabilize resonance contributor I (Fig. 1b); a positively biased field will destabilize it. In principle, a positively biased field should also lower the barrier height by stabilizing resonance contributor III. However, configuration III has much less inherent stability than contributor I, as the electronegative oxygen prefers to bear a negative rather than a positive charge. As a result, this configuration contributes only at strong positive fields, outside of the experimental range (see Supplementary Figs 3–8). Within the experimental range, our calculations predict that the frequency of adduct formation should increase systematically as the strength of the negatively biased field increases, up to a factor of 1.5 at  $-0.75$  V, while remaining relatively independent of EEF strength for positively biased fields.

To test these predictions experimentally, we attached the NB to the surface of a flat gold electrode, and the furan to the STM gold tip. We then undertook a series of STM break-junction approach (BJ) experiments known as ‘blinking’<sup>27,28</sup>. The blinking technique detects the formation of molecular bridges between an STM tip and a substrate electrode—while they are fixed at a specific electrode–electrode distance—by imposing an initial set-point tunnelling current (Fig. 3 and Supplementary Information, section 2). After the set-point current is reached, the feedback loop is turned off and the current is monitored. Current jumps (blinks) appear when a molecular bridge spans the gap between the electrodes (Fig. 3c). We detected blinks in conductance of magnitude  $(1.5\text{--}2) \times 10^{-6} G_0$  when the tip and substrate were separated by a distance that allows the Diels–Alder reaction to occur (about 1 nm). These junctions formed only when both reactants were present. When either reactant was removed from its respective electrode, or when their saturated analogues were used (2-methyl-3-tetrahydrofuranthiol on the tip, or a hydrogenated version of NB on the surface—a system that is structurally identical but which lacks the diene–dienophile character), there was no evidence of molecular-bridge formation (Supplementary Information, section 2). Hence, the junctions are formed through the Diels–Alder reaction. At positive voltage biases (surface positive), the frequency of molecular-bridge formation is constant

(five blinks per hour) over a wide range of applied biases. In contrast, at negative biases the frequency is clearly affected by the strength of the field, and increases from five blinks per hour at a bias of  $-0.05$  V, to 25 blinks per hour at a bias of  $-0.75$  V (Fig. 4). These trends are in complete qualitative agreement with the theoretical predictions in Fig. 2d. Quantitatively, there are differences, which may relate to the difference in the realms being studied experimentally and computationally (single-molecule reaction rates versus bulk reaction rates), and/or the use of pre-complexes in calculating field effects on barrier heights (see Supplementary Information).

As further validation that the experiment is detecting the formation of carbon–carbon bonds, we note that the average lifetime of the blinks was 0.4 s (Fig. 3d), with poor dependence on the electric-field magnitude for positive biases up to 0.75 V (Supplementary Information, section 2). This lifetime is around the same as that observed for standard single-molecule wires that are thiolated at both ends<sup>6,24</sup>. A bias value higher than  $+0.75$  V or lower than  $-0.75$  V led to a drop in the lifetime of the junctions, owing to the instability of the gold–sulfur contacts. Hence, we kept the upper limit of the bias within the range  $+0.75$  V to  $-0.75$  V, to allow us to compare different biases while maintaining similar junction stabilities. We also confirmed the formation of mechanically stable carbon–carbon bonds by collecting pulling curves during the blinking events (Supplementary Figs 2–6), or by performing pushing/pulling cycles (Supplementary Figs 2–7). Pulling curves collected over the blinks showed a plateau with an average pulling length of 0.2 nm, which corresponds to the stretching of the single-molecule bridge. When the pulling was exerted over the tunnelling background or on random noise (Supplementary Figs 2–6), a clean exponential decay was observed, testifying that the above blinks are attributable to a stable molecular junction, rather than resulting from the migration of gold atoms or from fluctuations in molecular conformations<sup>28,29</sup>.

We further confirmed that robust molecular junctions were formed using an STM-BJ approach<sup>5,30</sup> referred to as tapping. Here, the furan-modified tip was repeatedly driven into and out of contact with the NB-modified substrate (Supplementary Figs 2–7): when the reactants were mechanically brought together, molecular junctions formed; when the tip was pulled away, the junctions broke. During the pulling portion of the cycles, we detected plateaus in the current-versus-distance curves of the same conductance magnitude as that observed in the blinking experiments ( $(1.5\text{--}2) \times 10^{-6} G_0$ ; Supplementary Information, section 2 and Supplementary Figs 2–7), further supporting the idea that stable molecular wires form when the two reactants are brought together under an electric field. Moreover, changes to the magnitude and direction of the field applied across the reactants affected the rate of the Diels–Alder reaction in a similar manner to that found via



**Figure 4 | Frequency of blinks (junctions) as a function of the applied bias.** Positive and negative biases are plotted in red and blue, respectively. To keep the distance between the surface and the tip constant across the bias range, we used the same set-point current and changed the bias while the STM feedback was turned off. We performed blinking experiments over periods of one hour. At the end of each one-hour period, we changed the furan-modified STM tip and its lateral position with respect to the surface to compensate for the loss of reactants. We repeated this procedure eight times, giving each bias point (magnitude and direction) a total experimental time of eight hours. We changed the chronology of the selected biases randomly for each repeat. Error bars represent the standard deviation from the eight (one-hour) intervals. The dashed lines are included for visual guidance.

blinking. Tapping data show that product formation increases by 4.4-fold, from 4.2% (252 product molecules out of 6,000 attempts) to 18.6% (1,116 product molecules out of 6,000 attempts) when the surface bias is changed from  $-0.05$  V to  $-0.75$  V (Supplementary Figs 2–7).

We have presented the first (to our knowledge) experimental evidence of a non-redox, bond-forming process being accelerated by an oriented EEF. Our experimental results are qualitatively consistent with theoretical calculations, and result from the ability of the electric field to electrostatically stabilize a minor charge-separated resonance contributor of the transition state. This ability to manipulate chemical reactions with electric fields offers proof-of-principle for a change in our approach to heterogeneous catalysis.

Received 16 May 2015; accepted 7 January 2016.

- Meir, R., Chen, H., Lai, W. & Shaik, S. Oriented electric fields accelerate Diels–Alder reactions and control the endo/exo selectivity. *ChemPhysChem* **11**, 301–310 (2010).
- Gryn'ova, G., Marshall, D. L., Blanksby, S. J. & Coote, M. L. Switching radical stability by pH-induced orbital conversion. *Nature Chem.* **5**, 474–481 (2013).
- Gryn'ova, G. & Coote, M. L. Origin and scope of long-range stabilizing interactions and associated SOMO–HOMO conversion in distonic radical anions. *J. Am. Chem. Soc.* **135**, 15392–15403 (2013).
- Shaik, S., de Visser, S. P. & Kumar, D. External electric field will control the selectivity of enzymatic-like bond activations. *J. Am. Chem. Soc.* **126**, 11746–11749 (2004).
- Xu, B. & Tao, N. J. Measurement of single-molecule resistance by repeated formation of molecular junctions. *Science* **301**, 1221–1223 (2003).
- Haiss, W. *et al.* Measurement of single molecule conductivity using the spontaneous formation of molecular wires. *Phys. Chem. Chem. Phys.* **6**, 4330–4337 (2004).
- Haiss, W. *et al.* Precision control of single-molecule electrical junctions. *Nature Mater.* **5**, 995–1002 (2006).
- Sini, G., Maitre, P., Hiberty, P. C. & Shaik, S. S. Covalent, ionic and resonating single bonds. *J. Mol. Struct. Theochem* **229**, 163–188 (1991).
- Shaik, S., Danovich, D., Wu, W. & Hiberty, P. C. Charge-shift bonding and its manifestations in chemistry. *Nature Chem.* **1**, 443–449 (2009).
- Shurki, A., Hiberty, P. C. & Shaik, S. Charge-shift bonding in group IVB halides: a valence bond study of  $MH_3-Cl$  ( $M = C, Si, Ge, Sn, Pb$ ) molecules. *J. Am. Chem. Soc.* **121**, 822–834 (1999).
- Fischer, H. & Radom, L. Factors controlling the addition of carbon-centered radicals to alkenes—an experimental and theoretical perspective. *Angew. Chem. Int. Edn* **40**, 1340–1371 (2001).

- Lai, W., Li, C., Chen, H. & Shaik, S. Hydrogen-abstraction reactivity patterns from A to Y: the valence bond way. *Angew. Chem. Int. Edn* **51**, 5556–5578 (2012).
- Warshel, A. *et al.* Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106**, 3210–3235 (2006).
- Hirao, H., Chen, H., Carvajal, M. A., Wang, Y. & Shaik, S. Effect of external electric fields on the C–H bond activation reactivity of nonheme iron–oxo reagents. *J. Am. Chem. Soc.* **130**, 3319–3327 (2008).
- Lai, W., Chen, H., Cho, K.-B. & Shaik, S. External electric field can control the catalytic cycle of cytochrome P450cam: a QM/MM Study. *J. Phys. Chem. Lett.* **1**, 2082–2087 (2010).
- Fried, S. D., Bagchi, S. & Boxer, S. G. Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science* **346**, 1510–1514 (2014).
- Klinska, M., Smith, L. M., Gryn'ova, G., Banwell, M. G. & Coote, M. L. Experimental demonstration of pH-dependent electrostatic catalysis of radical reactions. *Chem. Sci.* **6**, 5623–5627 (2015).
- Franchi, P., Mezzina, E. & Lucarini, M. SOMO–HOMO conversion in distonic radical anions: an experimental test in solution by EPR radical equilibration technique. *J. Am. Chem. Soc.* **136**, 1250–1252 (2014).
- Aleman, M. *et al.* Electric field-induced isomerization of azobenzene by STM. *J. Am. Chem. Soc.* **128**, 14446–14447 (2006).
- Gorin, C. F., Beh, E. S. & Kanan, M. W. An electric field-induced change in the selectivity of a metal oxide-catalyzed epoxide rearrangement. *J. Am. Chem. Soc.* **134**, 186–189 (2012).
- Nicolaou, K. C., Snyder, S. A., Montagnon, T. & Vassilikogiannakis, G. The Diels–Alder reaction in total synthesis. *Angew. Chem. Int. Edn* **41**, 1668–1698 (2002).
- Paddon-Row, M. N. Investigating long-range electron-transfer processes with rigid, covalently linked donor–(norbornylogous bridge)–acceptor systems. *Acc. Chem. Res.* **27**, 18–25 (1994).
- Darwish, N., Paddon-Row, M. N. & Gooding, J. J. Surface-bound norbornylogous bridges as molecular rulers for investigating interfacial electrochemistry and as single molecule switches. *Acc. Chem. Res.* **47**, 385–395 (2014).
- Darwish, N. *et al.* Observation of electrochemically controlled quantum interference in a single anthraquinone-based norbornylogous bridge molecule. *Angew. Chem. Int. Edn* **51**, 3203–3206 (2012).
- Darwish, N. *et al.* Probing the effect of the solution environment around redox-active moieties using rigid anthraquinone terminated molecular rulers. *J. Am. Chem. Soc.* **134**, 18401–18409 (2012).
- Darwish, N. *et al.* Electroactive self-assembled monolayers of unique geometric structures by using rigid norbornylogous bridges. *Chemistry* **18**, 283–292 (2012).
- Diez-Pérez, I. *et al.* Rectification and stability of a single molecular diode with controlled orientation. *Nature Chem.* **1**, 635–641 (2009).
- Donhauser, Z. J. *et al.* Conductance switching in single molecules through conformational changes. *Science* **292**, 2303–2307 (2001).
- Claridge, S. A., Schwartz, J. J. & Weiss, P. S. Electrons, photons, and force: quantitative single-molecule measurements from physics to biology. *ACS Nano* **5**, 693–729 (2011).
- Tao, N. J. Electron transport in molecular junctions. *Nature Nanotechnol.* **1**, 173–181 (2006).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** This research was supported by the MINECO Spanish National Project (no. CTQ2012-36090) and an EU Reintegration Grant (FP7-PEOPLE-2010-RG-277182), and by resources provided at the NCI National Facility systems at the Australian National University through the National Computational Merit Allocation Scheme, supported by the Australian Government. N.D. thanks the European Union for a Marie Curie IIF Fellowship. I.D.-P. thanks the Ramon y Cajal program (MINECO, no. RYC-2011-07951) for financial support. S.C. thanks the University of Wollongong for a Vice Chancellor Fellowship, and the Australian National Fabrication Facility for financial support. A.C.A. thanks the Spanish Ministerio de Educación for an FPU fellowship. M.L.C. acknowledges financial support from the Australian Research Council (ARC) and an ARC Future Fellowship, and discussions with M. Banwell. We also acknowledge funding from the ARC Centre of Excellence Scheme (project no. CE 140100012).

**Author Contributions** A.C.A., N.D. and I.D.-P. carried out the STM experiments and analysed the data. N.J.B. and N.L.H. performed the quantum-chemical modelling, with input from M.L.C. S.C. carried out the synthetic work. All authors conceived the work, and designed and discussed the experiments. M.L.C. and S.C. wrote the manuscript, with substantial contributions from the other authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.L.C. (michelle.coote@anu.edu.au), I.D.-P. (isma\_diez@ub.edu), N.D. (ndarwish@ibecbarcelona.eu) or S.C. (sciampi@uow.edu.au).

# Upper-plate controls on co-seismic slip in the 2011 magnitude 9.0 Tohoku-oki earthquake

Dan Bassett<sup>1</sup>, David T. Sandwell<sup>1</sup>, Yuri Fialko<sup>1</sup> & Anthony B. Watts<sup>2</sup>

The March 2011 Tohoku-oki earthquake was only the second giant (moment magnitude  $M_w \geq 9.0$ ) earthquake to occur in the last 50 years and is the most recent to be recorded using modern geophysical techniques. Available data place high-resolution constraints on the kinematics of earthquake rupture<sup>1</sup>, which have challenged prior knowledge about how much a fault can slip in a single earthquake and the seismic potential of a partially coupled megathrust interface<sup>2</sup>. But it is not clear what physical or structural characteristics controlled either the rupture extent or the amplitude of slip in this earthquake. Here we use residual topography and gravity anomalies to constrain the geological structure of the overthrusting (upper) plate offshore northeast Japan. These data reveal an abrupt southwest–northeast-striking boundary in upper-plate structure, across which gravity modelling indicates a south-to-north increase in the density of rocks overlying the megathrust of 150–200 kilograms per cubic metre. We suggest that this boundary represents the offshore continuation of the Median Tectonic Line, which onshore juxtaposes geological terranes composed of granite batholiths (in the north) and accretionary complexes (in the south)<sup>3</sup>. The megathrust north of the Median Tectonic Line is interseismically locked<sup>2</sup>, has a history of large earthquakes (18 with  $M_w > 7$  since 1896) and produced peak slip exceeding 40 metres in the Tohoku-oki earthquake<sup>1</sup>. In contrast, the megathrust south of this boundary has higher rates of interseismic creep<sup>2</sup>, has not generated an earthquake with  $M_j > 7$  (local magnitude estimated by the Japan Meteorological Agency) since 1923, and experienced relatively minor (if any) co-seismic slip in 2011<sup>1</sup>. We propose that the structure and frictional properties of the overthrusting plate control megathrust coupling and seismogenic behaviour in northeast Japan.

The seismic moment of megathrust earthquakes is proportional to the product of rupture area, the average amount of slip, and the effective shear modulus<sup>4</sup>. Most variability is usually attributed to rupture area and so research has focused on the structural and geometrical barriers that segment plate boundaries and limit the dimensions, and thus magnitude, of earthquakes. Previous studies have shown the roughness of subducting plates to be one of the first-order controls on rupture dimensions<sup>5–7</sup>; but the amplitude of slip is proportional to the rupture area and the stress drop<sup>8</sup>, and may be harder to predict. In particular, of the five earthquakes since 1900 that had  $M_w \geq 9$ , the March 2011 Tohoku-oki earthquake had the smallest rupture area by 50% (that is, its rupture area was 0.5 that of the next-smallest earthquake in Kamchatka in 1952), but had the largest maximum displacements by 75% (that is, its maximum slip was 1.75 times that of the next-largest earthquake in Chile in 1960) (ref. 1 and Extended Data Fig. 1).

Geodetic observations show that most of the elastic strain energy released in megathrust earthquakes is accumulated in the overthrusting plate<sup>2</sup> so mapping along-strike variations in upper-plate structure is critical to understanding fault loading. But identifying upper-plate structural variations from the topography and gravity fields is difficult because the relatively small-amplitude, short-wavelength structure is

masked by the large-amplitude, trench-normal topography and gravity gradients associated with subduction zones. Here the application of spectral averaging routines designed to isolate and remove these gradients<sup>9,10</sup> provides observations that reveal how the crustal structure of the northeast Japan forearc influenced the rupture pattern of the Tohoku-oki earthquake.

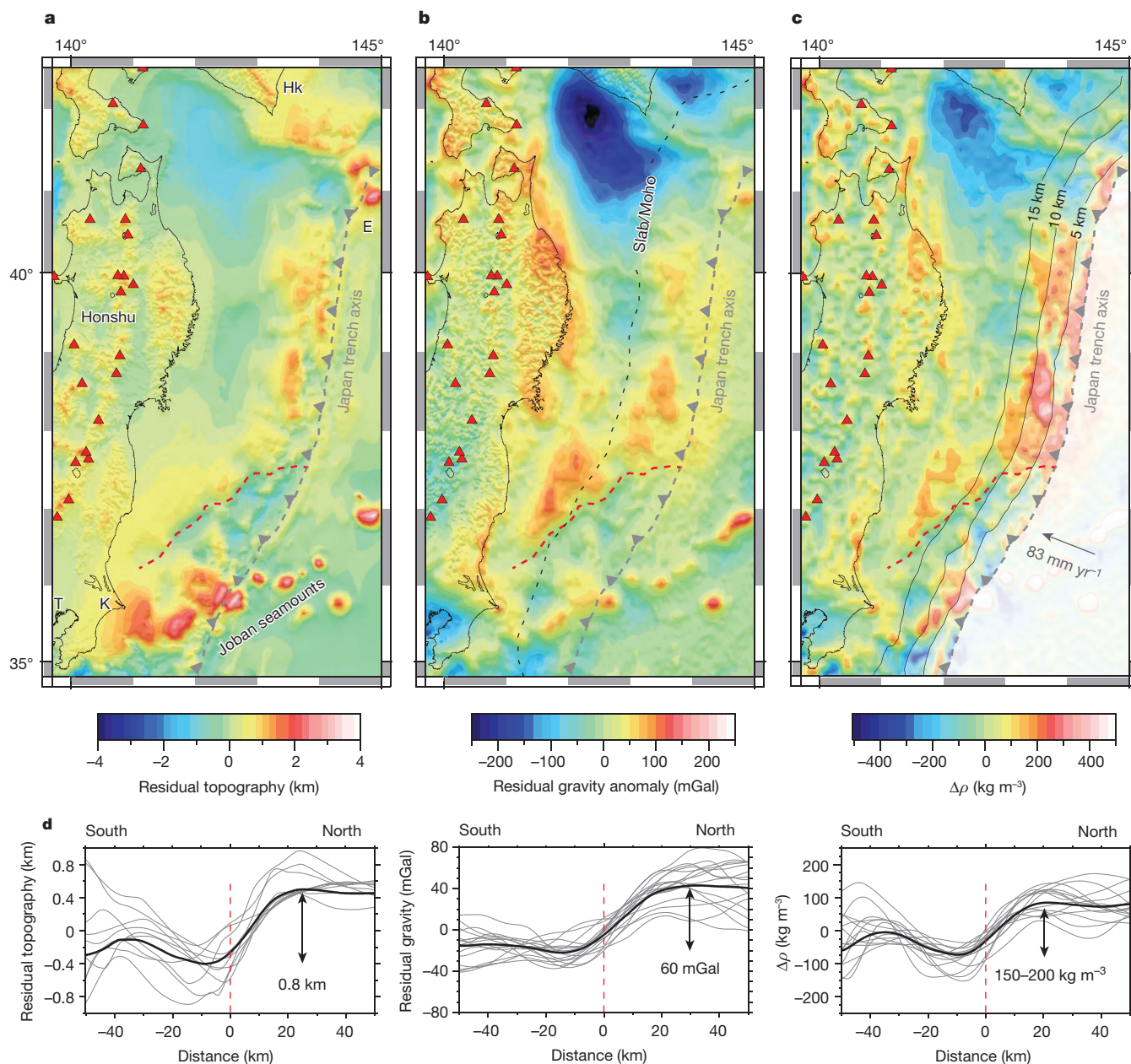
Residual topography and gravity anomalies were calculated by regionally subtracting a spectral average of the trench-normal topography and gravity anomalies, and are shown in Fig. 1 (see Methods and Extended Data Fig. 2 for grid processing). The Pacific oceanic plate is subducting beneath Honshu at 80–85 mm yr<sup>-1</sup> and the submarine volcanoes of Erimo and the Joban seamounts are observed as positive residual topographic anomalies with amplitude >2 km both seaward and landward of the trench axis. The absence of similar anomalies along-strike suggests that the northeast Japan megathrust is relatively smooth. Across the forearc, our analysis reveals an abrupt and approximately linear transition in structure striking southwest–northeast (red dashed line in Fig. 1). Residual topography and gravity anomalies regionally increase from south to north across this boundary by ~0.8 km and ~60 mGal respectively (Fig. 1d). A second across-forearc transition of similar bathymetric character is located southeast of Hokkaido. The smooth and homogeneous nature of the subducting Pacific plate seaward of the trench axis strongly suggests that both transitions are related to the structure of the overthrusting plate.

Extension of the southern Japan forearc boundary to the trench axis makes it unlikely that variations in forearc Moho depth contribute to the along-strike contrast in gravity anomalies. We thus attribute residual gravity anomalies to trench-slope topography and lateral variations in the density of rocks comprising the overthrusting plate. The density contrasts associated with these variations are estimated by discretizing the forearc into 10 km × 10 km vertical prisms extending between the sea floor and the seismically constrained base of the forearc crust (Extended Data Figs 3–7), with a single density anomaly calculated for each prism (see Methods). The contrast in residual gravity anomalies can be explained by a south-to-north increase in the mean density of forearc rocks by ~150–200 kg m<sup>-3</sup> (Fig. 1c, d). The inverse proportionality between density contrasts and crustal thickness makes this a minimum estimate and the reduction in crustal thickness near the trench axis increases the magnitude of density contrasts (Fig. 1c). This analysis shows that the sharp change in trench-slope morphology is coincident with a transition in the density of the materials that comprise the southern Japan forearc crust.

The onshore geology<sup>11</sup> of Japan provides a physical interpretation of this transition (Fig. 2). The Median Tectonic Line (MTL) is the most prominent structural boundary and clearly separates two contrasting geological terrains. On the continent (north) side, a ~20-km-thick granitic upper-crust reflects intrusion of arc melt since the Cambrian Period<sup>3</sup>. In contrast, on the ocean (south) side, the forearc crust is entirely composed of variably metamorphosed Late Mesozoic to Cenozoic accretionary complexes<sup>3</sup>. A variety of tectonic models have been proposed to explain the unnatural juxtaposition of these terrains<sup>3</sup>,

<sup>1</sup>Scripps Institution of Oceanography, La Jolla, California 92093, USA. <sup>2</sup>Department of Earth Sciences, University of Oxford, South Parks Road, Oxford, UK.





**Figure 1 | Forearc anomalies in the northeast Japan subduction zone.** **a**, Residual topography. Dashed red and grey lines mark the forearc segment boundary and the trench-axis respectively. Triangles show arc volcanoes. Hk, Hokkaido; E, Erimo seamount; T, Tokyo; K, Kashima. **b**, Residual gravity anomalies. The black dashed line marks the intersection of the subducting slab with the forearc Moho. **c**, Mean forearc density

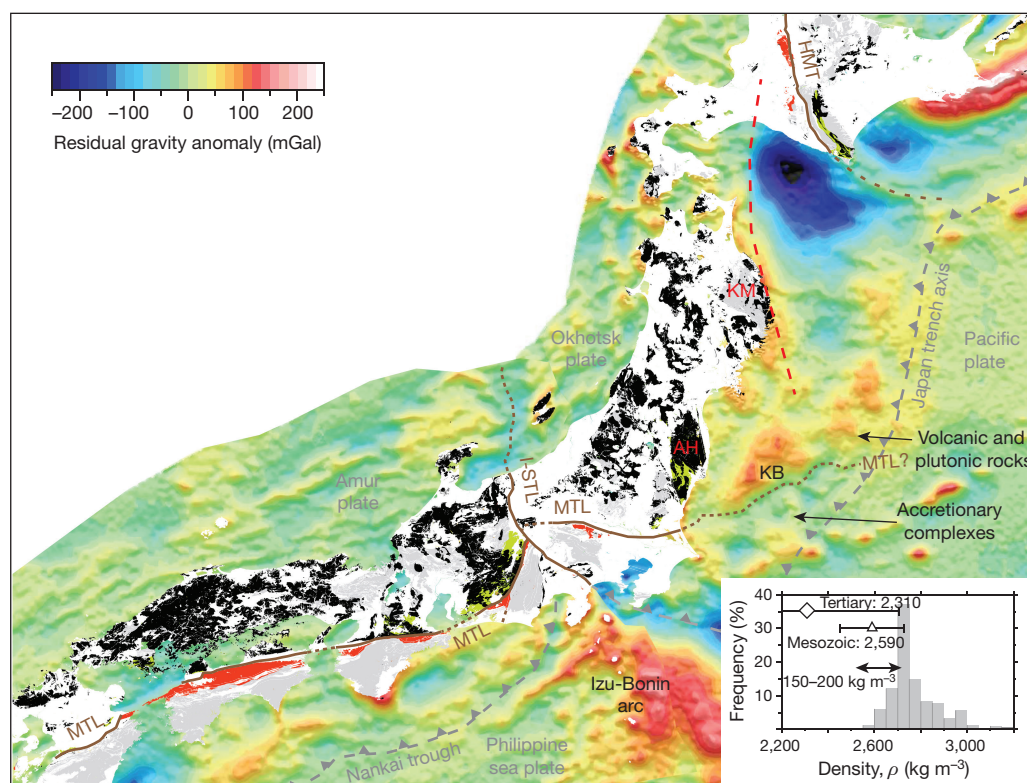
anomalies. Contours (5-km increment) show forearc crustal thickness. **d**, Profiles (grey lines) perpendicularly traversing the forearc segment boundary showing the south-to-north increases in topography ( $\sim 0.8$  km), residual gravity anomaly ( $\sim 60$  mGal) and mean forearc density anomalies ( $\sim 150$ – $200$  kg m $^{-3}$ ). The mean for each ensemble of profiles is plotted in black.

but magnetotelluric<sup>12</sup> and active source seismic<sup>13,14</sup> data show a shallow ( $35^{\circ}$ – $45^{\circ}$ ) north-dipping fault geometry that is most consistent with a thrust faulting origin either along a low-angle mid-crustal detachment<sup>3</sup> or the Cretaceous subduction megathrust<sup>14</sup>. These data further demonstrate that the fundamental contrast in upper-plate structure expressed in surface geology across the MTL persists to depths of at least 20 km (ref. 13).

We propose that the MTL extends offshore Kashima, connecting to the step in residual gravity and bathymetry. The implication is that the abrupt change in forearc structure represents the lithological juxtaposition of granitic batholiths to the north and accretionary complexes to the south. This suggestion is consistent with the offshore extent of positive aeromagnetic anomalies, which have been modelled as batholiths 10–15 km thick<sup>15</sup>. The trend of aeromagnetic anomalies (red dashed line in Fig. 2) correlates the largest-amplitude residual gravity anomaly

(labelled KB) with the Early Cretaceous Kitakami batholith, which is exposed onshore in the Kitakami mountains (KM) (Fig. 2). The presence of low-to-mid-pressure metamorphic rocks within the southern region of the Abukuma highland (AH) is consistent with the MTL geometry inferred offshore.

The most interesting and important aspect of this forearc transition is that it is highly correlated with the seismogenic behaviour of the megathrust, as shown in Fig. 3. First, the forearc segment boundary is associated with a sharp north-to-south reduction in the number of intermediate-magnitude earthquakes (Fig. 3a). Second, historical rupture areas of large ( $M_w/1 > 7.0$ ) megathrust earthquakes between 1896 and March 2011 (Supplementary Table 2) are similarly focused north of the proposed extension of the MTL (Fig. 3b); and the megathrust to the south, where the forearc is characterized by negative bathymetric and gravimetric anomalies, has not generated an earthquake with



**Figure 2 | Simplified geology<sup>11</sup> and major tectonic boundaries<sup>3</sup> of Japan.** Volcanic and plutonic rocks (Cambrian to middle Miocene) are shown in black. Accretionary complexes (Jurassic to middle Miocene) are shown in grey. Metamorphic rocks (Cretaceous to middle Miocene) are shown in red and green. Main faults are shown in brown. MTL geometry is constrained by the juxtaposition of high-pressure footwall rocks (red shading indicates the Sanbagawa and Shimanto accretionary complexes) with coeval low-pressure granitic hanging-wall rocks (green shading indicates the Ryoke-Sanyo belt). Note the similar transition in peak metamorphic pressure observed across the Hidaka main thrust (HMT)

in Hokkaido. KM, Kitakami mountains; KB, Kitakami batholith; AH, Aburatsubo highland; I-STL, Itoigawa-Shizuoka tectonic line. The strike of the Kitakami aeromagnetic anomalies is shown as a dashed red line<sup>15</sup>. The inset compares the density distribution of granitic rocks east of the volcanic arc in northeastern Honshu<sup>30</sup> and the mean dry densities for Tertiary and Mesozoic sedimentary rocks (error bars show one standard deviation)<sup>31</sup>. The modelled density contrast across the forearc segment boundary ( $\sim 150\text{--}200\text{ kg m}^{-3}$ ) is within the range expected across the MTL.

$M_j \geq 7.0$  in the >90-year duration of the Japan Meteorological Agency (<http://www.jma.go.jp/en/quake/>) earthquake catalogue. Third, there is a strong correlation between forearc structure and the distribution of co-seismic slip in the 2011  $M_w$  9.0 Tohoku-oki earthquake<sup>1</sup> (Fig. 3c). The Tohoku earthquake filled a seismic gap as defined by the rupture areas of earlier large thrust earthquakes (Fig. 3b) and most of the moment release occurred north of the forearc segment boundary in regions characterized by positive residual gravity anomalies. This correlation appears to be robust, as shown by a comparison with different published co-seismic slip models of the Tohoku earthquake that employed different data types and inversion algorithms (Extended Data Fig. 8). Finally, geodetically constrained interseismic deformation models all show a high degree of fault locking within the region of positive residual gravity<sup>2</sup>, but creep to the south of the forearc segment boundary (Fig. 3d, Supplementary Fig. 2).

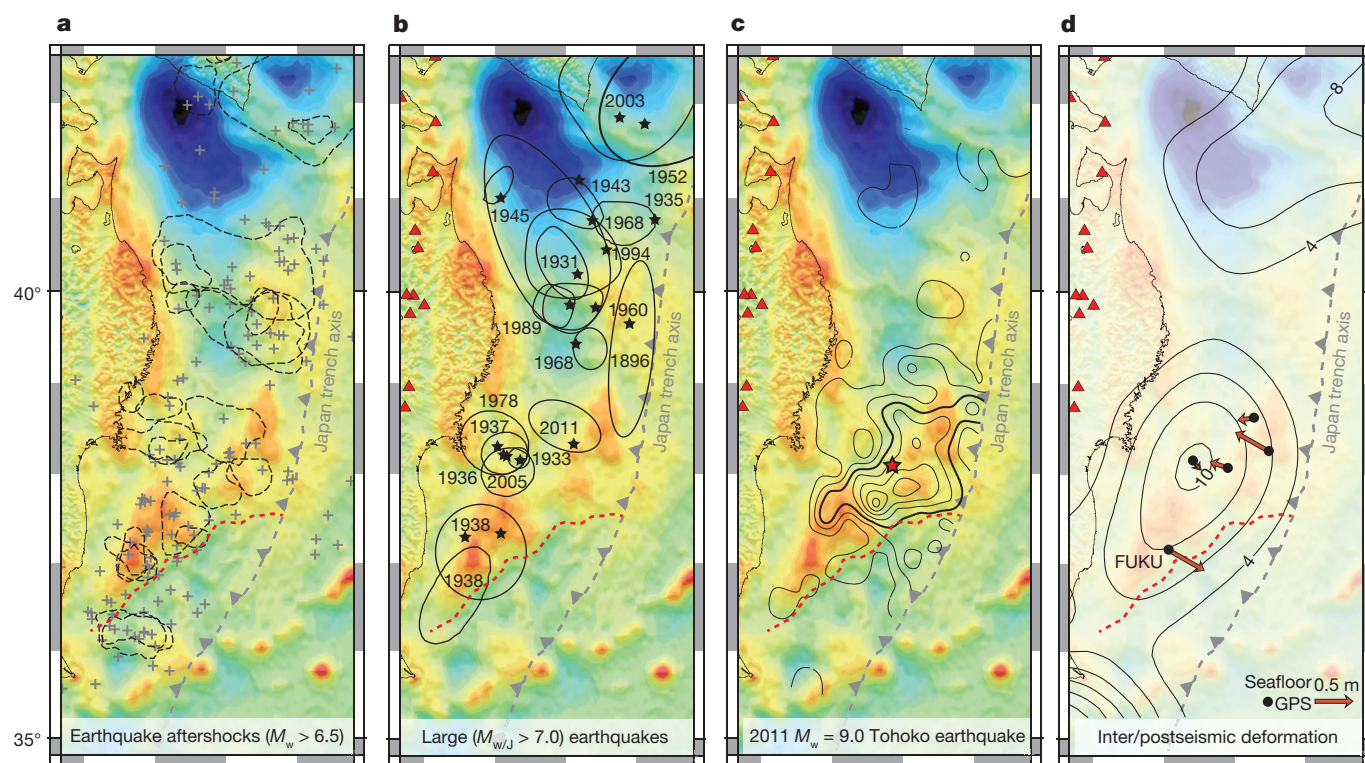
The simplest interpretation of the relationships described above is that upper-plate lithology modulates the frictional behaviour of the megathrust. Frictional properties of typical crystalline rocks comprising the volcanic arc are characterized by a relatively high static coefficient of friction ( $\mu_0 = 0.6\text{--}0.8$ ) and strong velocity weakening<sup>16,17</sup>. Increasing clay content decreases the coefficient of friction ( $\mu_0 = 0.2\text{--}0.4$  in gouge with >50 wt% clay)<sup>16,18</sup> and promotes a more stable response to perturbations in fault slip rate<sup>18</sup>. Materials south of the forearc segment boundary are thus expected to be both weaker and less velocity weakening.

Numerical models of earthquake cycles incorporating rate- and state-dependent friction show that in the interseismic period, asperities are loaded at a rate modulated by their strength and location

relative to other asperities, and the flexural rigidity of the upper plate, which controls the extent of stress shadowing<sup>19,20</sup>. Stress increases are maximum on asperities with the highest  $\mu_0$  and the most negative values of the rate-dependent friction parameter  $a - b$ , relative to adjacent fault regions ( $a$  and  $b$  represent the magnitudes of the direct and evolution effect in friction respectively)<sup>21</sup>. Correspondingly, the stress drop and amplitude of co-seismic slip are amplified by sharper contrasts in frictional properties at the asperity boundaries<sup>22</sup>. Dynamic rupture simulations also show that dynamic rupture fronts decelerate as they penetrate into unloaded, velocity-strengthening, or compliant (less rigid) regions, which may ultimately arrest co-seismic ruptures<sup>23</sup>.

Under this framework, the anomalous nature of the Tohoku asperity may be attributed to the sharp transition in  $\mu_0$  and/or  $a - b$  across the MTL. The abrupt nature of this transition results in the local development of a large stress concentration in the interseismic period due to creep in (presumably) velocity-strengthening regions down-dip and south of the asperity. Low seismicity north and south of the asperity (Fig. 3a, b) may reflect reduced interseismic stressing rates within the Tohoku 'stress shadow'. Simulations of earthquake cycles on a fault with heterogeneous distributions of  $\mu_0$  and  $a - b$  further reveal that large events can occur within the area that occasionally ruptures in great earthquakes (Extended Data Fig. 10). We propose that stress heterogeneities either side of the MTL may explain both the small slip area and the large slip amplitude of the Tohoku-oki earthquake rupture. (1) The extent of co-seismic slip was limited by the inability of the rupture front to penetrate into the low-stress or velocity-strengthening fault segment south of the MTL. (2) The highly stressed or strongly





**Figure 3 | Slip behaviour of the northeast Japan megathrust.**

**a**, Instrumental earthquake record. Grey plus symbols show the epicentres of earthquakes in the JMA catalogue (1923–2015) with  $M_I \geq 6.5$ . Dashed ellipses show the aftershock area of thrust earthquakes in the Global Centroid Moment Tensor (<http://www.globalcmt.org>) catalogue (period 1976–2014) with  $6.5 \leq M_w < 8$ . No aftershock areas cross the forearc segment boundary (dashed red line). **b**, Rupture areas for large ( $M_{w/J} > 7.0$ ) megathrust earthquakes between 1896 and the 2011

Tohoku-oki earthquake (Supplementary Table 3). **c**, Coseismic slip contours (10-m increment) for the 2011  $M_w$  9.0 Tohoku-oki earthquake<sup>1</sup>. The 20-m slip contour (thicker contour line) defines the Tohoku asperity. **d**, Inter/postseismic deformation. Contours show interseismic back-slip rate (increment  $2 \text{ cm yr}^{-1}$ )<sup>2</sup>. Arrows show 1-year postseismic displacements of sea-floor GPS sites<sup>24</sup>. The fast seaward motion of site FUKU is associated with shallow afterslip<sup>24</sup>.

velocity-weakening area with high rates of tectonic loading north of the MTL resulted in the large slip amplitude<sup>1</sup>. This interpretation implies accelerated postseismic afterslip within the ‘stress shadow’ areas of the Tohoku asperity. Robust afterslip has indeed been inferred in that area from the fast seaward motion of the sea-floor Global Positioning System (GPS) site FUKU<sup>24</sup> (Fig. 3d and Extended Data Fig. 9). The seaward motion of this site is contrary to the postseismic landward motions observed at sea-floor GPS sites within the main rupture area, and is probably caused by substantial afterslip south of the forearc segment boundary<sup>24</sup>. Regional sea-floor geodetic measurements from a cabled seismological and geodetic observatory<sup>25</sup> will place important constraints on the extent of post-seismic creep and moment accumulation rates, and provide further insights into tectonic loading of the Tohoku asperity, fault-slip behaviour south of the MTL and, ultimately, seismic hazard in Kashima and eastern Tokyo.

The correlation between the forearc segment boundary and megathrust slip behaviour appears to persist west of the intersection between the slab with the forearc Moho. This may reflect subduction of eroded upper-plate materials beyond the forearc Moho, thereby extending the influence of upper-plate lithology on the frictional properties of the megathrust to greater depth. The obliquity of the MTL may cause subduction of eroded materials to reduce the along-strike gradient in frictional properties, but both the location and magnitude of this effect will depend on the depth range of tectonic erosion. Finally, the contrast in mean forearc density and trench-slope topography across the MTL will result in an along-strike variation in lithostatic pressure of 40–60 MPa. Although considerably lower than dip-parallel variations in lithostatic stress associated with slab-dip and trench-slope topography, dip-parallel lithostatic gradients occur both north and south of the MTL and the along-strike contrast in lithostatic

stress may still be important in influencing the lateral distribution of plate locking at a given depth.

Two additional insights from this study may be important for understanding seismic hazard in other subduction zones. The first comes from the observation that the accreted terranes that appear to be creeping south of the forearc segment boundary<sup>2</sup> also overlie the Nankai megathrust in southwest Japan, which is characterized by interseismic locking<sup>26</sup> and produced large earthquakes in 1944 ( $M_w$  8.1), 1946 ( $M_w$  8.3) and 1968 ( $M_w$  7.5)<sup>27</sup>. On one hand, this may reflect differences in the relative proportion of subducting and overthrusting plate materials within the megathrust shear zone, and the strong dependence of the  $a - b$  value of clay-rich sediments on lithologic heterogeneity<sup>28</sup>. On the other hand, the amplitude of stress heterogeneities is dependent on spatial gradients in  $\mu_0$  and  $a - b$ . The apparent low seismicity north and south of the Tohoku asperity might be due to ‘stress shadowing’, resulting in interseismic stressing rates on the megathrust even lower than those implied by models of interseismic coupling (Fig. 3d). In Nankai, the composition of the overthrusting plate is more homogeneous along-strike, which may result in smaller variations in fault frictional properties, and more uniform interseismic loading rates. These factors are likely to encourage earthquake ruptures in Nankai to extend over larger areas. The second insight is that low residual gravity anomalies are not a good indicator of seismogenic behaviour, as proposed in previous studies<sup>27,29</sup>. The residual topography and gravity fields are now available for all subduction zones on Earth<sup>9,10</sup>, but their real utility comes from enabling existing and future seismological and geodetic observations to be considered in the context of subducting and overthrusting plate structure. This study shows that these considerations are an essential component of hazard assessment.



**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 26 June; accepted 11 December 2015.**

- Minson, S. *et al.* Bayesian inversion for finite fault earthquake source models—II: the 2011 great Tohoku-oki, Japan earthquake. *Geophys. J. Int.* **198**, 922–940 (2014).
- Suwa, Y., Miura, S., Hasegawa, A., Sato, T. & Tachibana, K. Interplate coupling beneath NE Japan inferred from three-dimensional displacement field. *J. Geophys. Res.* **111**, B04402 (2006).
- Isozaki, Y., Aoki, K., Nakama, T. & Yanai, S. New insight into a subduction-related orogen: a reappraisal of the geotectonic framework and evolution of the Japanese Islands. *Gondwana Res.* **18**, 82–105 (2010).
- Aki, K. Estimation of earthquake moment, released energy, and stress-strain drop from G-wave spectrum. *Bull. Earthq. Res. Inst.* **44**, 23–88 (1966).
- Wang, K. & Bilek, S. L. Fault creep caused by subduction of rough seafloor relief. *Tectonophysics* **610**, 1–24 (2014).
- Robinson, D. P., Das, S. & Watts, A. B. Earthquake rupture stalled by a subducting fracture zone. *Science* **312**, 1203–1205 (2006).
- Sparkes, R., Tilmann, F., Hovius, N. & Hillier, J. Subducted seafloor relief stops rupture in South American great earthquakes: implications for rupture behaviour in the 2010 Maule, Chile earthquake. *Earth Planet. Sci. Lett.* **298**, 89–94 (2010).
- Knopoff, L. Energy release in earthquakes. *Geophys. J. Int.* **1**, 44–52 (1958).
- Bassett, D. & Watts, A. B. Gravity anomalies, crustal structure, and seismicity at subduction zones: 1. Seafloor roughness and subducting relief. *Geochem. Geophys. Geosyst.* **16**, 1508–1540 (2015).
- Bassett, D. & Watts, A. B. Gravity anomalies, crustal structure, and seismicity at subduction zones: 2. Interrelationships between fore-arc structure and seismogenic behavior. *Geochem. Geophys. Geosyst.* **16**, 1541–1576 (2015).
- Seamless Digital Geological Map of Japan 1: 200,000. Jul 3, 2012 version, Research Information Database DB084 ([https://gbank.gsj.jp/seamless/index\\_en.html?](https://gbank.gsj.jp/seamless/index_en.html?)), Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology (2012).
- Goto, T., Yamaguchi, S., Sumitomo, N. & Yaskawa, K. The electrical structure across the Median Tectonic Line in east Shikoku, southwest Japan. *Earth Planets Space* **50**, 405–415 (1998).
- Kawamura, T., Onishi, M., Kurashimo, E., Ikawa, T. & Ito, T. Deep seismic reflection experiment using a dense receiver and sparse shot technique for imaging the deep structure of the Median Tectonic Line (MTL) in east Shikoku, Japan. *Earth Planets Space* **55**, 549–557 (2003).
- Sato, H., Kato, N., Abe, S., Van Horne, A. & Takeda, T. Reactivation of an old plate interface as a strike-slip fault in a slip-partitioned system: Median Tectonic Line, SW Japan. *Tectonophysics* **644**, 58–67 (2015).
- Finn, C. Aeromagnetic evidence for a buried Early Cretaceous magmatic arc, northeast Japan. *J. Geophys. Res.* **99**, 22165–22185 (1994).
- Morrow, C., Moore, D. E. & Lockner, D. The effect of mineral bond strength and adsorbed water on fault gouge frictional strength. *Geophys. Res. Lett.* **27**, 815–818 (2000).
- Mitchell, E., Fialko, Y. & Brown, K. Temperature dependence of frictional healing of Westerly granite: experimental observations and numerical simulations. *Geochem. Geophys. Geosyst.* **14**, 567–582 (2013).
- Numelin, T., Marone, C. & Kirby, E. Frictional properties of natural fault gouge from a low-angle normal fault, Panamint Valley, California. *Tectonics* **26** (2), 1–14 (2007).
- Kanamori, H. The nature of seismicity patterns before large earthquakes. *Earthquake Prediction* **4**, 1–19 (1981).
- Hetland, E. & Simons, M. Post-seismic and interseismic fault creep. II: Transient creep and interseismic stress shadows on megathrusts. *Geophys. J. Int.* **181**, 99–112 (2010).
- Kaneko, Y., Avouac, J.-P. & Lapusta, N. Towards inferring earthquake patterns from geodetic observations of interseismic coupling. *Nature Geosci.* **3**, 363–369 (2010).
- Hillers, G. & Wesnousky, S. Scaling relations of strike-slip earthquakes with different slip-rate-dependent properties at depth. *Bull. Seismol. Soc. Am.* **98**, 1085–1101 (2008).
- Tinti, E., Bizzarri, A. & Cocco, M. Modeling the dynamic rupture propagation on heterogeneous faults with rate- and state-dependent friction. *Ann. Geophys.* **48** (2), 327–345 (2005).
- Sun, T. & Wang, K. Viscoelastic relaxation following subduction earthquakes and its effects on afterslip determination. *J. Geophys. Res.* **120**, 1329–1344 (2015).
- Kanazawa, T. in *Underwater Technology Symposium IEEE International* 1–5, <http://dx.doi.org/10.1109/UT.2013.6519911> (IEEE, 2013).
- Wallace, L. M. *et al.* Enigmatic, highly active left-lateral shear zone in southwest Japan explained by aseismic ridge collision. *Geology* **37**, 143–146 (2009).
- Wells, R. E., Blakely, R. J., Sugiyama, Y., Scholl, D. W. & Dinterman, P. A. Basin-centered asperities in great subduction zone earthquakes: a link between slip, subsidence, and subduction erosion? *J. Geophys. Res.* **108** (B10), 1–30 (2003).
- den Hartog, S., Niemeijer, A. & Spiers, C. New constraints on megathrust slip stability under subduction zone P–T conditions. *Earth Planet. Sci. Lett.* **353/354**, 240–252 (2012).
- Song, T. & Simons, M. Large trench-parallel gravity variations predict seismogenic behavior in subduction zones. *Science* **301**, 630–633 (2003).
- Okuma, S. & Kanaya, H. Petrophysical Data Base of Basement Rocks in Japan for the 21st Century (PB-Rock 21). RIO-DB (Research Information Data Base) **87**, <http://riodb02.ibase.aist.go.jp/pb-rock21/index.html> (Geological Survey of Japan, AIST, 2011).
- Murata, Y., Suda, Y., Kikuchi, T. & Chōsajo, C. *Rock Physical Properties of Japan: Density, Magnetism, P-Wave Velocity, Porosity, Thermal Conductivity* (Geological Survey of Japan, 1991).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank S. Naif, C. Davies, C. Twardzik and S. Das for suggestions. Figures and grid processing was conducted using the Generic Mapping Tools (GMT) (<http://gmt.soest.hawaii.edu/>). D.B. was supported by a University of Oxford Clarendon Scholarship, by a Green Foundation Postdoctoral Fellowship in the Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, by the National Geospatial Agency (HMO1771310008), and by the Scripps Seafloor Electromagnetics Consortium (<http://marineemlab.ucsd.edu/semc.html>).

**Author Contributions** D.B. and A.B.W. conceived the study and conducted grid processing. D.B. and D.T.S. calculated density anomalies. Y.F. conducted numerical simulations of earthquake cycles. D.B., D.T.S. and Y.F. wrote the initial manuscript. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. ([dbassett@ucsd.edu](mailto:dbassett@ucsd.edu)).

## METHODS

**Spectral averaging of topography and gravity grids.** Residual anomalies are calculated using the satellite-derived free-air gravity anomaly grid of Sandwell *et al.*<sup>32</sup> and the GEBCO shipboard bathymetry grid<sup>33</sup>. Free-air gravity anomalies onshore are reduced to Bouguer gravity anomalies using a crustal density of  $2,670 \text{ kg m}^{-3}$ . Bathymetry and gravity grids are sampled by 1,200-km-long trench-normal profiles, which are centred on the bathymetrically defined trench-axis and spaced  $\sim 25 \text{ km}$  along strike. The mean cross-sectional structure of the subduction zone is calculated as the spectral average across each ensemble of profiles. The linearity of the fast Fourier transform and its inverse means that spectral averages should be indistinguishable from arithmetic averages, but small differences ( $\leq 10 \text{ m}$  and  $\leq 2 \text{ mGal}$ ) are present, resulting from the finite precision of the fast Fourier transform and associated down-sampling<sup>9</sup>. Maintaining the geometry of the trench axis, average profiles are extended along-strike to produce grids of each ensemble average profile, which is then subtracted from the original data set to produce grids of residual topography and residual gravity anomalies. In contrast to profile-based residual calculations, subtraction of an ensemble average grid preserves the full  $1 \text{ min} \times 1 \text{ min}$  resolution of the original data sets. This processing methodology has been applied globally<sup>9,10</sup> and is illustrated in Extended Data Fig. 2.

**Calculation of density anomalies.** Using the average topographic profile and active-source seismic constraints on across-arc crustal structure (Supplementary Table 2 and Extended Data Figs 4 and 5), the ensemble average gravity anomaly profile can be well fitted using realistic density distributions for the crust ( $\sim 2,800 \text{ kg m}^{-3}$ ) and mantle ( $\sim 3,100 \text{ kg m}^{-3}$ ) (Extended Data Fig. 3). This shows that the ensemble average gravity anomaly captures the broad crustal architecture of the subduction zone, enabling short-wavelength residual gravity anomalies landward of the trench axis to be interpreted as reflecting crustal structure of the forearc, arc and backarc.

The forearc segment boundary extends to the trench axis with no change in amplitude across the intersection of the subducting Pacific plate with the forearc Mohorovic boundary (Moho) (Fig. 1b). It is thus unlikely (and nearer the trench impossible) that variations in forearc Moho depth contribute to the along-strike contrast in gravity anomalies. Hence, we attribute residual gravity anomalies to trench-slope topography and lateral changes in density of the overthrusting forearc crust. The magnitude of these changes are calculated by discretizing the forearc into  $10 \text{ km} \times 10 \text{ km}$  vertical prisms. The top of each prism is constrained by the sea floor and the base is constrained by either the forearc Moho or the top of the subducting slab (whichever is shallower). The geometry of both interfaces is constrained by active-source wide-angle seismic models (Extended Data Figs 4 and 5 and Supplementary Table 2). The initial density contrast  $\Delta\rho$  for each prism is calculated directly from the amplitude of residual gravity anomalies  $R_{\text{grav}}$ :

$$\Delta\rho = \frac{R_{\text{grav}}}{2\pi Gh}$$

where  $h$  is the crustal thickness of the forearc and  $G$  is the gravitational constant. Synthetic gravity anomalies are calculated using the *Fatiando:Gravmag* module<sup>34</sup>, which calculates the gravity effect of three-dimensional rectangular prisms using the formula of Plouff<sup>35</sup>. Outstanding residual gravity anomalies are used to update  $\Delta\rho$  within each prism. After 12 iterations the root-mean-square misfit between the observed and calculated residual gravity anomalies is  $< 1 \text{ mGal}$  (inset to Extended Data Fig. 6b). The magnitude of density contrasts is inversely proportional to the vertical extent of the causative prism. Our methodology thus yields minimum estimates with density contrasts distributed throughout the full thickness of the forearc crust,  $h$ . As shown by the contours in Fig. 1c, the larger density contrasts near the trench axis predominantly reflect the reduction in forearc crustal thickness (Extended Data Fig. 3c) although they may also reflect, in part, contrasting resistances to near-trench deformation. The distribution of density anomalies is not strongly dependent on the observations used to constrain forearc crustal thickness and experimentation between passive (SLAB1.0<sup>36</sup> and earthquake tomographic<sup>37</sup>) versus active-source seismic (Supplementary Table 2) constraints on subducting slab and forearc Moho geometries yield differences in density anomalies of  $< 10 \text{ kg m}^{-3}$  (Extended Data Fig. 7). These differences are an order of magnitude smaller than the  $\sim 150 \text{ kg m}^{-3}$  contrast in mean forearc density calculated across the forearc segment boundary (Fig. 1c-d).

**Comparison with earthquake distributions.** Forearc structure is compared with the slip behaviour of the megathrust using published slip models and aftershock distributions for earthquakes with  $M_{\text{w}}/J > 7$  occurring since 1896. This catalogue is presented as Supplementary Table 3. For earthquakes with  $6.5 \leq M_{\text{w}} \leq 7.5$  between 1976 and 2015 (duration of instrumental records), rupture areas are estimated from the distribution of aftershocks in the International Seismological Center (ISC) (<http://www.isc.ac.uk/>) earthquake catalogue. Global analyses have shown that

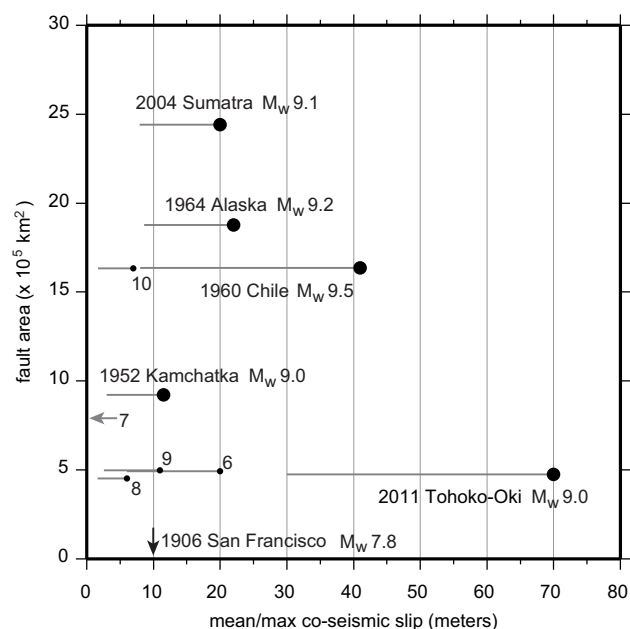
there is little change in aftershock areas after the first week<sup>38</sup>, and we estimate rupture areas from one-week aftershock distributions.

Tajima *et al.*<sup>39</sup> reviewed 44 published slip distributions for the Tohoku-oki earthquake. Most show the maximum slip to be  $\geq 40 \text{ m}$ . Following the asperity definition applied in northeast Japan by Yamanaka and Kikuchi<sup>40</sup>, we define the Tohoku asperity as regions where the co-seismic slip was  $\geq 20 \text{ m}$  and half the maximum slip amplitude (thickened contour Fig. 3c).

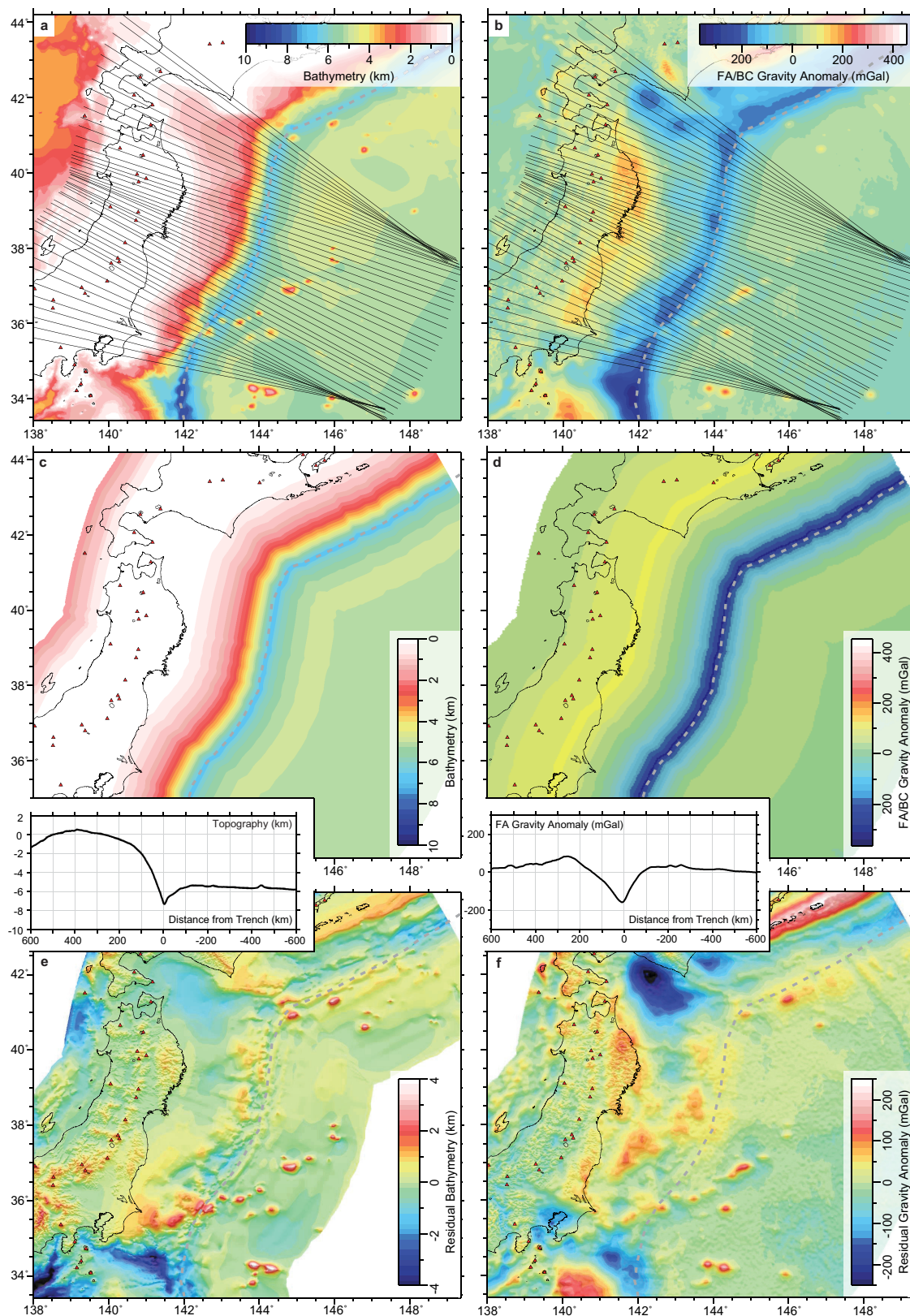
32. Sandwell, D. T., Müller, D. T., Smith, W. H. F., Garcia, E. & Francis, R. New global marine gravity from CryoSat-2 and Jason-1 reveals buried tectonic structure. *Science* **346**, 65–67 (2014).
33. Intergovernmental Oceanographic Commission, International Hydrographic Organisation, and British Oceanographic Data Centre. *Centenary Edition of the GEBCO Digital Atlas [CD-ROM]* ([http://www.gebco.net/data\\_and\\_products/gridded\\_bathymetry\\_data/gebco\\_one\\_minute\\_grid/](http://www.gebco.net/data_and_products/gridded_bathymetry_data/gebco_one_minute_grid/)) (British Oceanographic Data Centre, 2003).
34. Uieda, L., Oliveira, V. C., Jr & Barbosa, V. C. Modeling the Earth with Fatiando a Terra. *Proc. 12th 'Python in Science' Conf.* 96–103, <http://www.fatiando.org> (2013).
35. Plouff, D. Gravity and magnetic fields of polygonal prisms and application to magnetic terrain corrections. *Geophysics* **41**, 727–741 (1976).
36. Hayes, G. P., Wald, D. J. & Johnson, R. L. Slab1.0: a three-dimensional model of global subduction zone geometries. *J. Geophys. Res.* **117**, 1–15 (2012).
37. Katsumata, A. Depth of the Moho discontinuity beneath the Japanese islands estimated by traveltimes analysis. *J. Geophys. Res.* **115**, 1–17 (2010).
38. Henry, C. & Das, S. Aftershock zones of large shallow earthquakes: fault dimensions, aftershock area expansion and scaling relations. *Geophys. J. Int.* **147**, 272–293 (2001).
39. Tajima, F., Mori, J. & Kennett, B. L. A review of the 2011 Tohoku-Oki earthquake (Mw 9.0): large-scale rupture across heterogeneous plate coupling. *Tectonophysics* **586**, 15–34 (2013).
40. Yamanaka, Y. & Kikuchi, M. Asperity map along the subduction zone in northeastern Japan inferred from regional seismic data. *J. Geophys. Res.* **109**, 1–16 (2004).
41. Barrientos, S. E. & Ward, S. N. The 1960 Chile earthquake: inversion for slip distribution from surface deformation. *Geophys. J. Int.* **103**, 589–598 (1990).
42. Johnson, J. M., Satake, K., Holdahl, S. R. & Sauber, J. The 1964 Prince William Sound earthquake: joint inversion of tsunami and geodetic data. *J. Geophys. Res.* **101**, 523–532 (1996).
43. Chlieh, M. *et al.* Coseismic slip and afterslip of the great Mw 9.15 Sumatra–Andaman earthquake of 2004. *Bull. Seismol. Soc. Am.* **97**, S152–S173 (2007).
44. Johnson, J. M. & Satake, K. in *Seismogenic and Tsunamigenic Processes in Shallow Subduction Zones* 541–553 (Springer, 1999).
45. Delouis, B., Nocquet, J. M. & Vallée, M. Slip distribution of the February 27, 2010 Mw = 8.8 Maule earthquake, central Chile, from static and high-rate GPS, InSAR, and broadband teleseismic data. *Geophys. Res. Lett.* **37**, 1–7 (2010).
46. Kelleher, J., Savino, J., Rowlett, H. & McCann, W. Why and where great thrust earthquakes occur along island arcs. *J. Geophys. Res.* **79**, 4889–4899 (1974).
47. Beck, S. L. & Christensen, D. H. Rupture process of the February 4, 1965, Rat Islands earthquake. *J. Geophys. Res.* **96**, 2205–2221 (1991).
48. Johnson, J. M. *et al.* The 1957 great Aleutian earthquake. *Pure Appl. Geophys.* **142**, 3–28 (1994).
49. Ito, A. *et al.* Fault plane geometry in the source region of the 1994 Sanriku-oki earthquake. *Earth Planet. Sci. Lett.* **223**, 163–175 (2004).
50. Takahashi, N. *et al.* Seismic structure and seismogenesis off Sanriku region, northeastern Japan. *Geophys. J. Int.* **159**, 129–145 (2004).
51. Miura, S. *et al.* Structural characteristics off Miyagi forearc region, the Japan Trench seismogenic zone, deduced from a wide-angle reflection and refraction study. *Tectonophysics* **407**, 165–188 (2005).
52. Ito, A. *et al.* Bending of the subducting oceanic plate and its implication for rupture propagation of large interplate earthquakes off Miyagi, Japan, in the Japan Trench subduction zone. *Geophys. Res. Lett.* **32**, 1–4 (2005).
53. Miura, S. *et al.* Structural characteristics controlling the seismicity crustal structure of southern Japan Trench fore-arc region, revealed by ocean bottom seismographic data. *Tectonophysics* **363**, 79–102 (2003).
54. Fujie, G. *et al.* Along-trench structural variation and seismic coupling in the northern Japan subduction zone. *Earth Planets Space* **65**, 75–83 (2013).
55. Yamamoto, Y., Hino, R. & Shinohara, M. Mantle wedge structure in the Miyagi Prefecture forearc region, central northeastern Japan arc, and its relation to corner-flow pattern and interplate coupling. *J. Geophys. Res.* **116**, 1–18 (2011).
56. Fujie, G. *et al.* A significant relation between seismic activities and reflection intensities in the Japan Trench region. *Geophys. Res. Lett.* **29**, 1–4 (2002).
57. Wang, Z. & Zhao, D. Seismic imaging of the entire arc of Tohoku and Hokkaido in Japan using P-wave, S-wave and sP depth-phase data. *Phys. Earth Planet. Inter.* **152**, 144–162 (2005).
58. Simons, M. *et al.* The 2011 magnitude 9.0 Tohoku-Oki earthquake: mosaicking the megathrust from seconds to centuries. *Science* **332**, 1421–1425 (2011).
59. Ammon, C. J., Lay, T., Kanamori, H. & Cleveland, M. A rupture model of the 2011 off the Pacific coast of Tohoku earthquake. *Earth Planets Space* **63**, 693–696 (2011).
60. Yue, H. & Lay, T. Inversion of high-rate (1 sps) GPS data for rupture process of the 11 March 2011 Tohoku earthquake (Mw 9.1). *Geophys. Res. Lett.* **38**, 1–6 (2011).

61. Melgar, D. & Bock, Y. Kinematic earthquake source inversion and tsunami inundation prediction with regional geophysical data. *J. Geophys. Res.* **120**, 3324–3349 (2015).
62. Sato, T., Hiratsuka, S. & Mori, J. Precursory seismic activity surrounding the high-slip patches of the 2011 Mw 9.0 Tohoku-Oki earthquake. *Bull. Seismol. Soc. Am.* **103**, 3104–3114 (2013).
63. Ozawa, S. *et al.* Preceding, coseismic, and postseismic slips of the 2011 Tohoku earthquake, Japan. *J. Geophys. Res.* **117**, 1–20 (2012).
64. Fujii, Y. & Satake, K., Sakai, S. i., Shinohara, M. & Kanazawa, T. Tsunami source of the 2011 off the Pacific coast of Tohoku Earthquake. *Earth Planets Space* **63**, 815–820 (2011).
65. Asano, Y. *et al.* Spatial distribution and focal mechanisms of aftershocks of the 2011 off the Pacific coast of Tohoku Earthquake. *Earth Planets Space* **63**, 669–673 (2011).
66. Lapusta, N. & Rice, J. R., Ben-Zion, Y. & Zheng, G. Elastodynamic analysis for slow tectonic loading with spontaneous rupture episodes on faults with rate-and state-dependent friction. *J. Geophys. Res.* **105**, 23765–23789 (2000).
67. Kaneko, Y., Fialko, Y., Sandwell, D., Tong, X. & Furuya, M. Interseismic deformation and creep along the central section of the North Anatolian fault (Turkey): InSAR observations and implications for rate-and-state friction properties. *J. Geophys. Res.* **118**, 316–331 (2013).



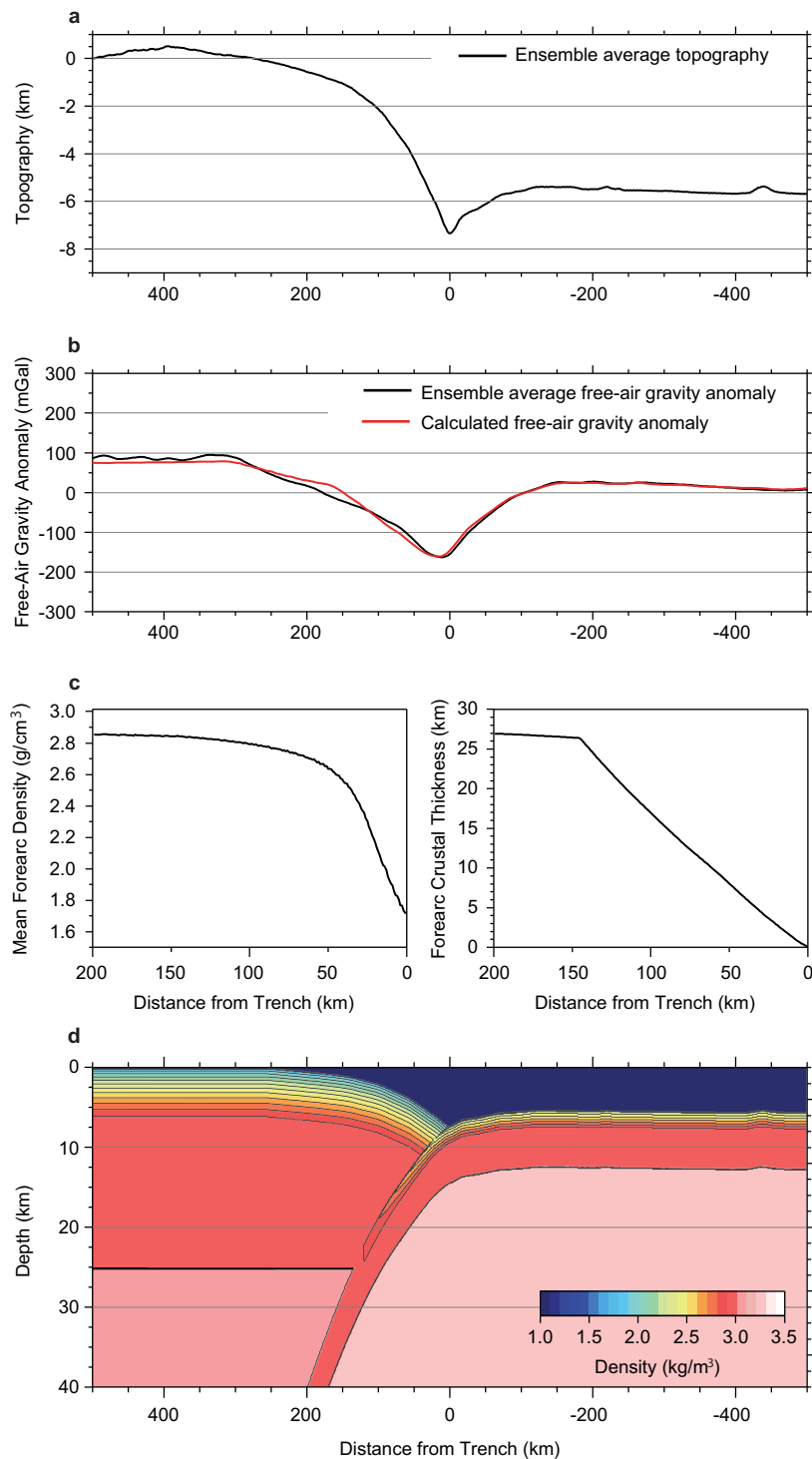


**Extended Data Figure 1 | Rupture areas and co-seismic slip amplitudes in Earth's largest earthquakes<sup>1,41-48</sup>.** Grey bars and black dots plot the mean and maximum amounts of co-seismic slip against the area of fault rupture. The five earthquakes with  $M_w \geq 9$  are labelled and numbers 6–10 refer to the location of smaller-magnitude earthquakes in the catalogue shown in Supplementary Table 1. Note the anomalously large ( $>70$  m) amount of co-seismic slip in the 2011 Tohoku-oki event<sup>1</sup>. The black arrow shows the maximum amount of slip in the 1906 San Francisco earthquake. The rupture area in this strike-slip event was  $6 \times 10^3$  km<sup>2</sup> and two orders of magnitude smaller than the megathrust rupture areas plotted here.



**Extended Data Figure 2 | Grid processing methodology.** Panels illustrate the ensemble-averaging and grid-processing methodology as applied at the northeast Japan subduction zone. Regional grids of bathymetry<sup>33</sup> and free-air/Bouguer-corrected (FA/BC) gravity anomaly<sup>32</sup> are sampled along trench-normal profiles (a and b). The spectral average is calculated from

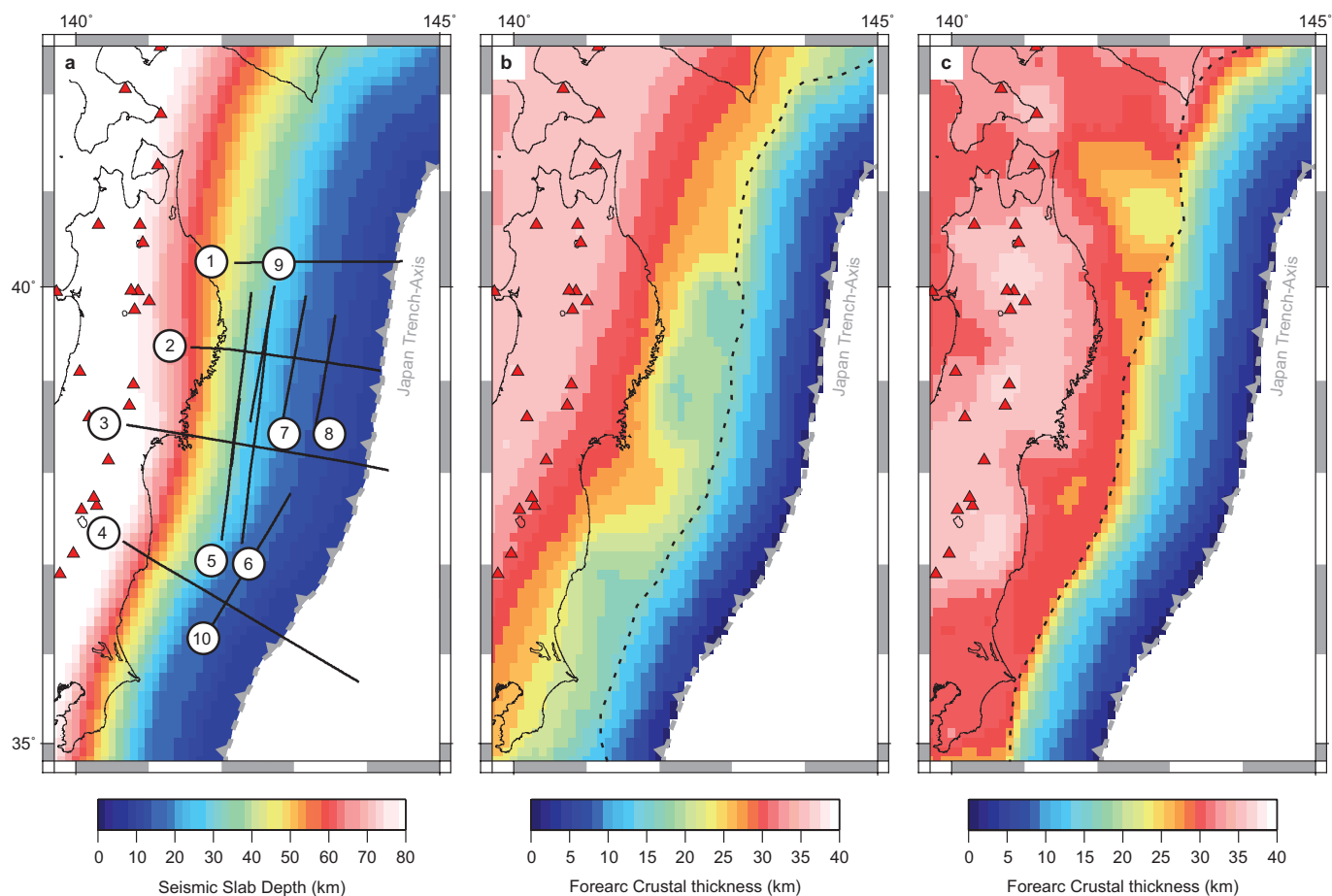
each ensemble of profiles (shown as insets to c and d). Maintaining the geometry of the trench, grids of the average profile are constructed (c and d), and subtracted from the original data sets to reveal residual bathymetry (e) and residual gravity anomalies (f). This technique of spectral averaging is identical to that applied globally<sup>9,10</sup>.



**Extended Data Figure 3 | Ensemble average profiles and two-dimensional gravity model.** **a**, Ensemble average topographic profile. **b**, Ensemble average gravimetric profile (black). The red profile shows the gravity anomaly calculated for the two-dimensional density structure shown in **d**. **c**, Mean forearc density (left) and forearc crustal thickness (right) plotted against distance from the trench axis. The larger amplitudes of density anomalies near the trench predominantly reflect the reduction in forearc crustal thickness  $h$  although they may also reflect, in part, contrasting resistances to near-trench deformation. **d**, Model of crustal structure for the northeast Japan subduction zone. This model is constructed using the mean geometry of the trench slope (shown in **a**)

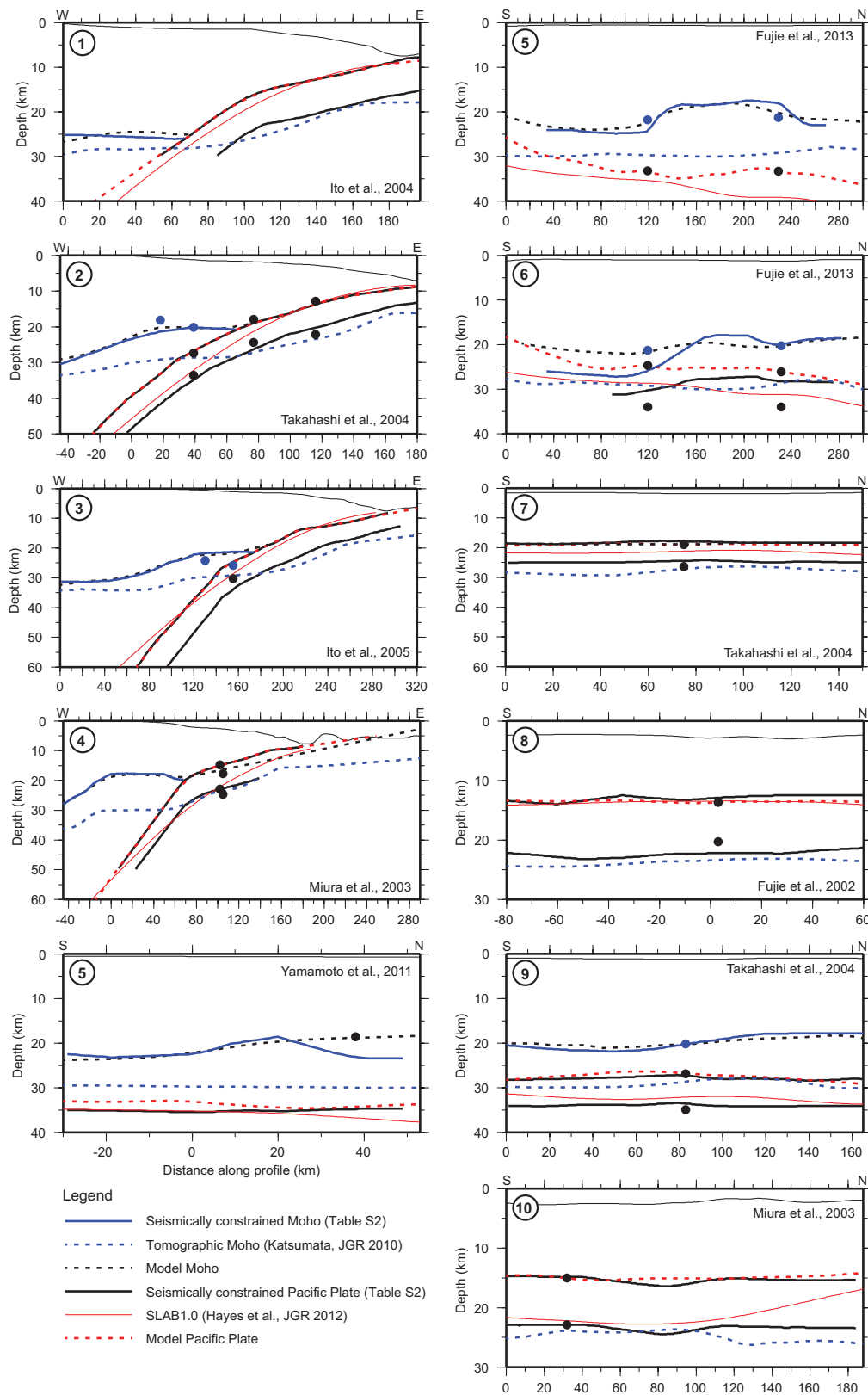
and subducting slab, seismic constraints on forearc and subducting slab ( $\sim 7$  km) crustal thicknesses (Supplementary Table 2), and using reasonable values for crustal ( $\sim 2,800 \text{ kg m}^{-3}$ ) and mantle densities ( $\sim 3,100 \text{ kg m}^{-3}$ ). The good fit observed in **b** between the ensemble average (black) and calculated (red) gravity anomalies shows that the ensemble average gravity anomaly captures the broad crustal architecture of the subduction zone, which enables the residual anomalies revealed following the removal of this average to be interpreted. The short-wavelength nature of residuals in the northeast Japan forearc suggests that most are related to crustal structure.





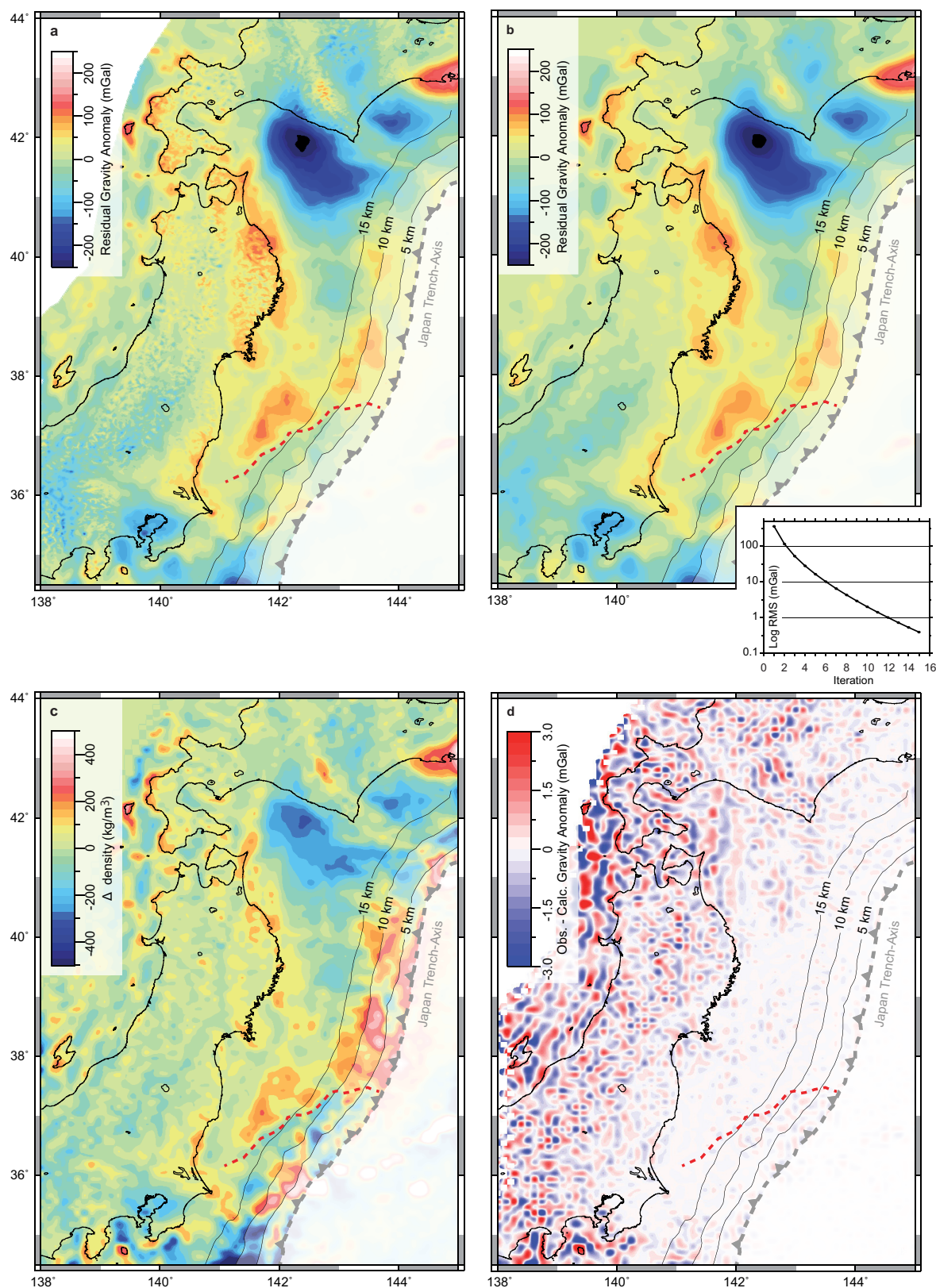
**Extended Data Figure 4 | Subducting slab geometry and forearc crustal thickness.** **a**, Geometry of the subducting Pacific Plate as constrained by linearly interpolating along-strike between active-source wide-angle seismic profiles<sup>49–56</sup>. Profiles are numbered as listed in Supplementary Table 2 and plotted in Extended Data Fig. 5. Red triangles show arc volcanoes. **b**, Forearc crustal thickness as constrained by the wide-angle

profiles shown in **a**. The dotted line marks the intersection of the subducting slab with the forearc Moho. Crustal thickness is calculated by subtracting the observed bathymetry from the seismically constrained base of the forearc crust. **c**, As in **b**, but with forearc Moho depth constrained by the tomographic model of Katsumata<sup>37</sup>.



**Extended Data Figure 5 | Wide-angle seismic models.** Profiles are numbered as in Extended Data Fig. 4a and Supplementary Table 2<sup>49–56</sup>. See legend for figure nomenclature. Dots show slab and Moho positions at profile intersections. Horizontal axes show model kilometres. The slab and Moho geometries shown for profiles 5 and 6 are from reflector

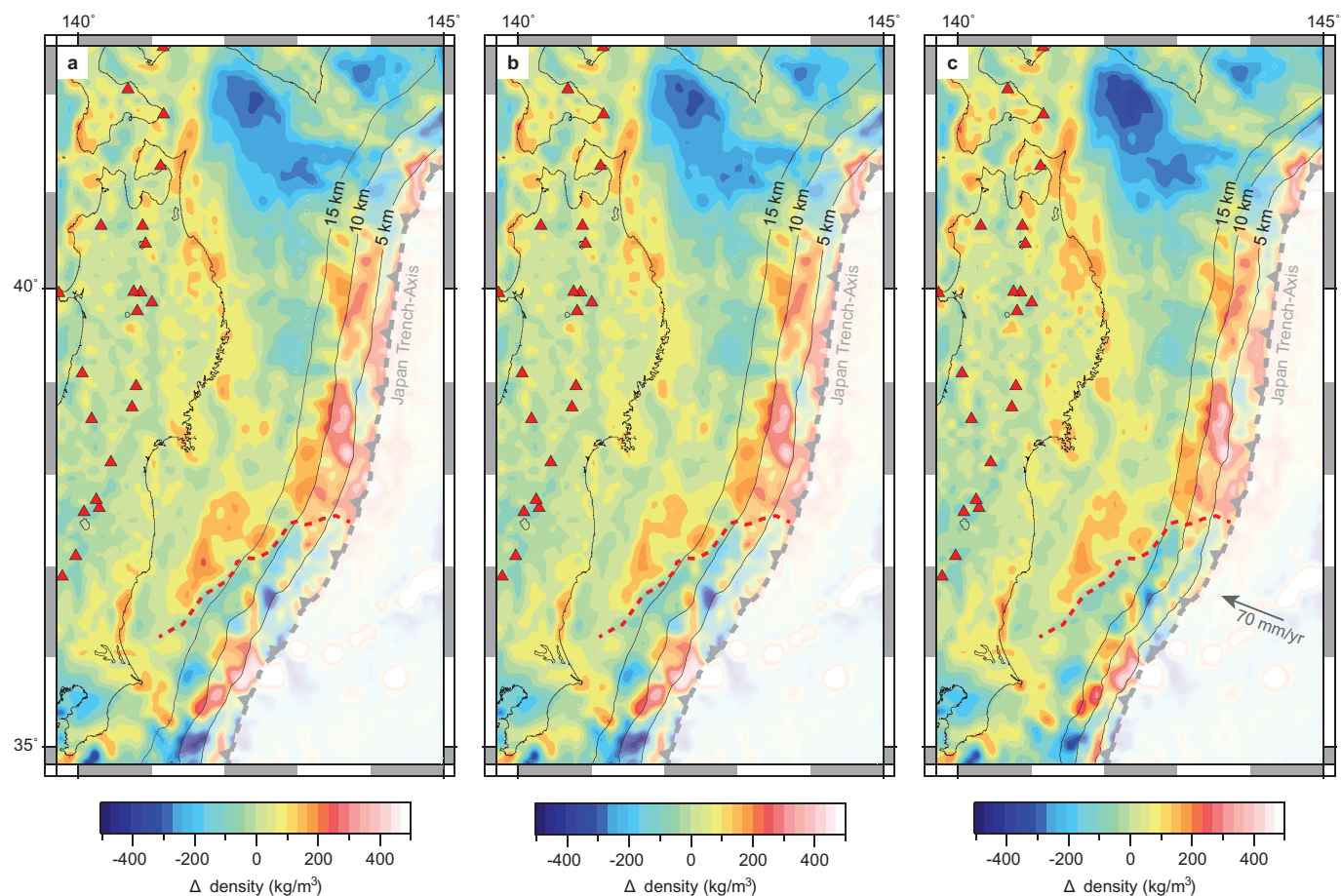
distributions imaged by travel-time mapping<sup>54</sup>. Intersecting profile 3 suggests that Moho reflectors interpreted south of 120 km on profile 6 may originate from the top of the subducting Pacific plate and forearc Moho constraints are only incorporated north of model kilometre 150.



**Extended Data Figure 6 | Calculation of forearc density anomalies.** **a**, Observed residual gravity anomalies. Black contours show forearc crustal thickness (5 km increment). **b**, Synthetic gravity anomalies calculated from the distribution of forearc density anomalies shown in **c**. **c**, Distribution of density anomalies. Density contrasts are constant within 10 km × 10 km vertical prisms extending between the seabed and either the top of the subducting slab or the forearc Moho (whichever is

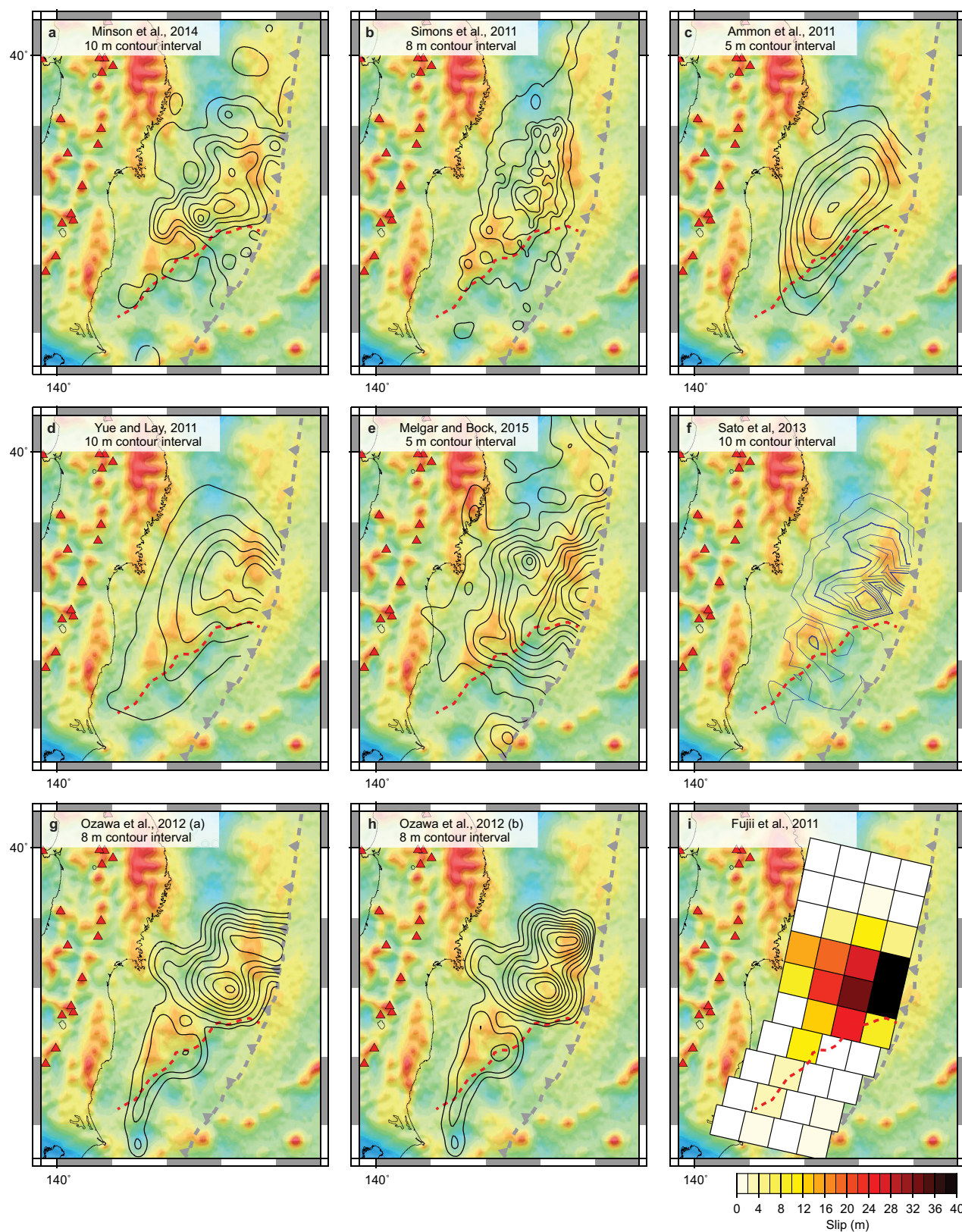
shallower). Initial density contrasts for each prism are estimated directly from residual gravity anomalies using the known thicknesses of each prism, with the difference between observed and synthetic residual gravity anomalies similarly applied to update model parameters. The inset to **b** shows the reduction in root-mean-square misfit with each update of model parameters. After 12 iterations the root-mean-square misfit < 1 mGal. **d**, Difference between observed and synthetic residual gravity anomalies.





**Extended Data Figure 7 | Density anomalies calculated using different constraints on forearc crustal thickness.** **a**, Density anomalies calculated using active-source seismic constraints on the slab and forearc Moho (Extended Data Fig. 4a and b). **b**, Density anomalies calculated using active-source seismic constraints on the geometry of the subducting Pacific Plate, but using the tomographic model of Katsumata<sup>37</sup> to constrain the forearc Moho (Extended Data Fig. 4c). **c**, Density anomalies calculated using SLAB1.0<sup>36</sup> for the subducting Pacific Plate and assuming a planar forearc Moho at the mean depth (25 km) determined by active-source

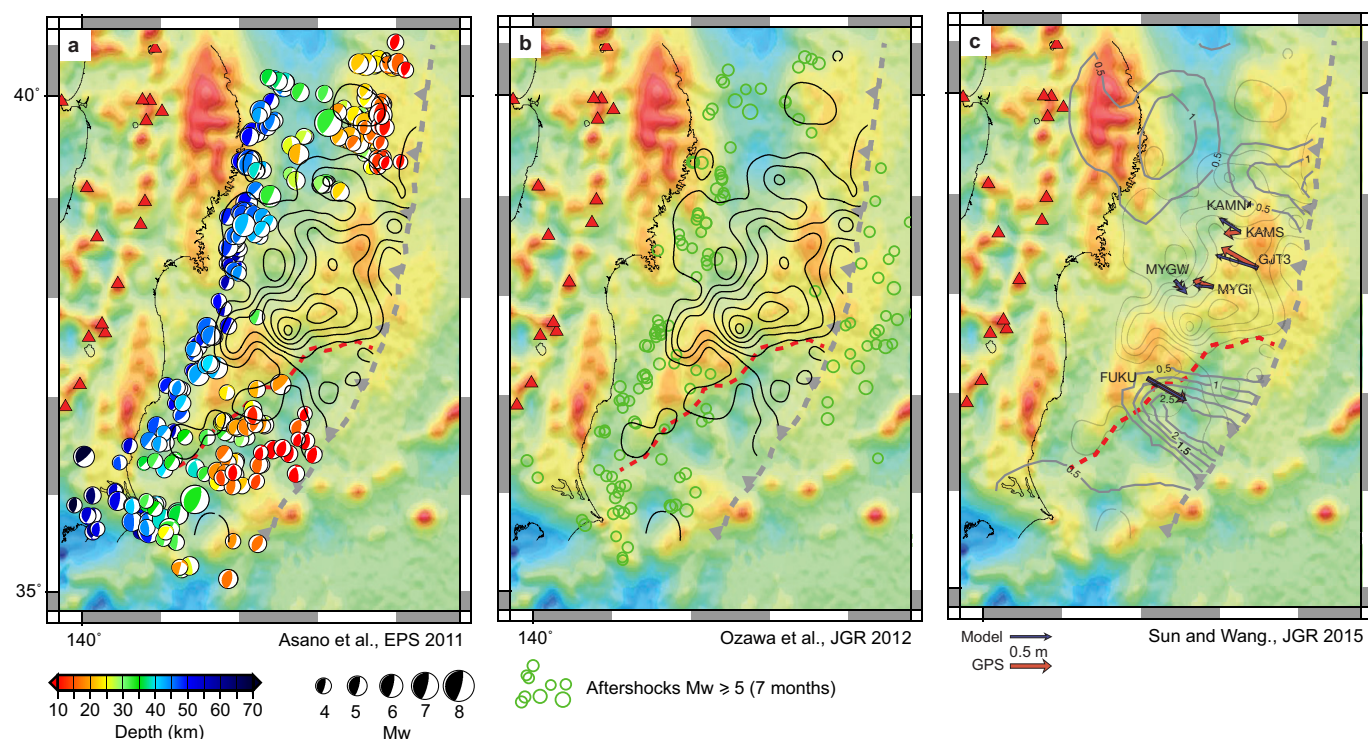
seismic models. The difference in density models calculated using these different model parameterizations are of the order of  $10\text{--}20\text{ kg m}^{-3}$ . All panels show a clear north-to-south reduction in density anomalies across the forearc segment boundary (red dashed line), and our interpretation of this contrast is not dependent on the observations used to constrain forearc crustal thickness. A comparison between forearc structure inferred from residual gravity anomalies and the seismic velocity structure of the forearc<sup>57</sup> is shown in Supplementary Fig. 1.



**Extended Data Figure 8 | Co-seismic slip models for the March 2011 Tohoku-oki earthquake.** Plots showing the correlation between overthrusting plate structure as constrained by residual gravity anomalies and the distribution of slip in the Tohoku-oki earthquake. These models have been constructed using different data types and inversion strategies. In all plots, grey and red dashed lines mark the trench axis and the forearc segment boundary respectively. Contour intervals are labelled and the outermost contour is 0. **a**, Minson *et al.*<sup>1</sup>; **b**, Simons, *et al.*<sup>58</sup>; **c**, Ammon *et al.*<sup>59</sup>; **d**, Yue and Lay<sup>60</sup>; **e**, Melgar and Bock<sup>61</sup>; **f**, Sato *et al.*<sup>62</sup>;

**g**, Ozawa *et al.*<sup>63</sup> (allowing slip at trench); **h**, Ozawa *et al.*<sup>63</sup> (imposing no-slip condition at trench); **i**, Fujii *et al.*<sup>64</sup>. In all plots, large co-seismic slip is focused north of the forearc segment boundary in regions characterized by positive residual gravity anomalies. Most models also show a sharp reduction in the magnitude of slip from north to south across the forearc segment boundary. The wide range of data types and inversion strategies represented by this ensemble of models suggests that the common features identified above are probably robust characteristics of the Tohoku-oki earthquake rupture.

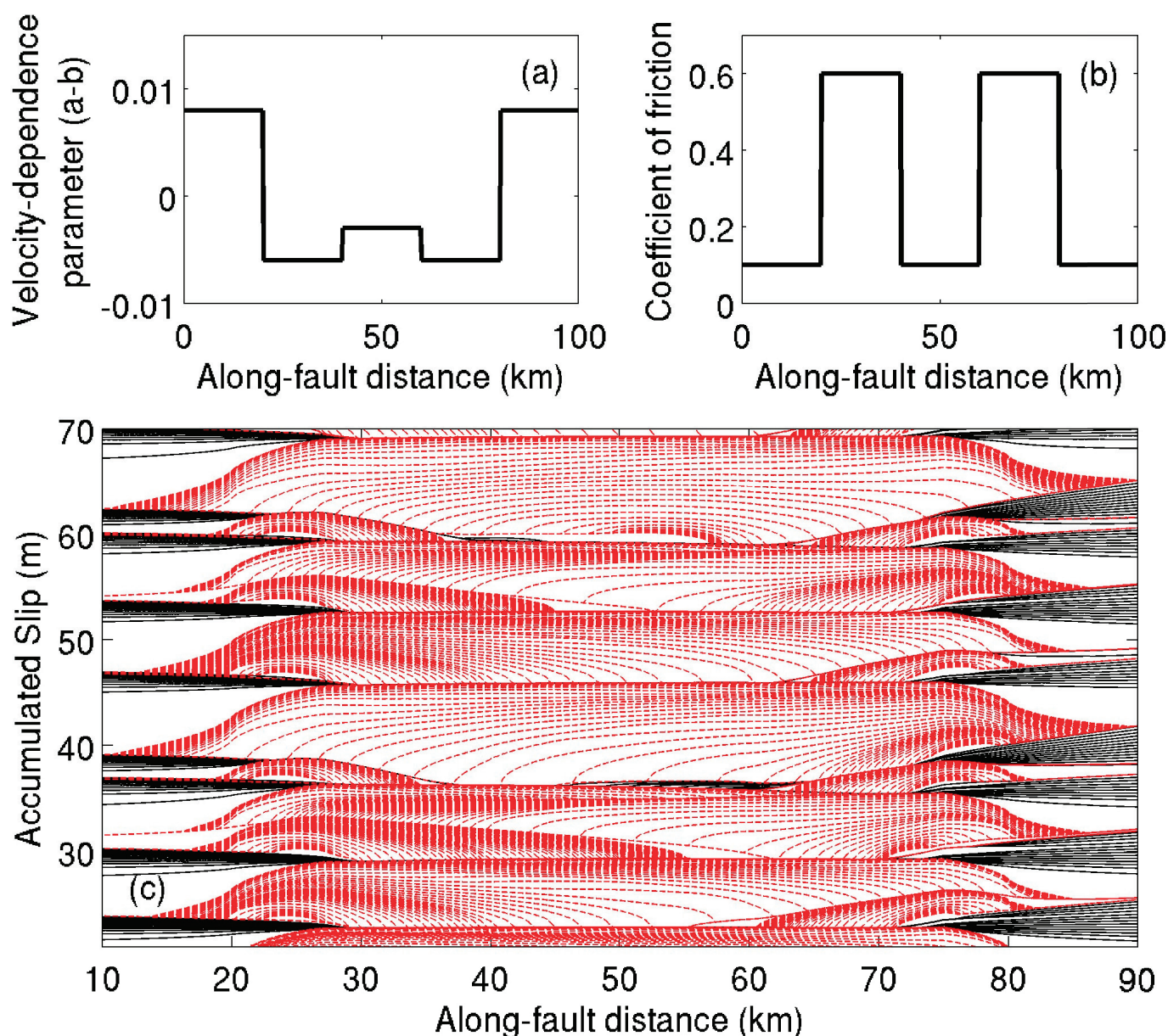




**Extended Data Figure 9 | Postseismic observations.** **a**, Aftershocks between March 11 and May 24 occurring on the subduction interface<sup>65</sup>. All plots show the trench axis (grey dashed line), forearc segment boundary (red dashed line) and contours (10 m) of co-seismic slip<sup>1</sup>. **b**, All aftershocks (variable location/mechanism) occurring within seven months of the Tohoku mainshock with  $M_w \geq 5$  (ref. 63). Panels **a** and **b** show that interplate aftershocks for the Tohoku-oki earthquake did not occur in areas that experienced large co-seismic displacements. The vast majority occur in regions surrounding the mainshock rupture area, the distribution of which supports the Bayesian slip distribution of ref. 1, and provides a useful constraint on the along-strike extent of co-seismic rupture. The negative correlation between co-seismic slip and aftershock locations is strongest for the aftershock locations of ref. 65, because they have a lower magnitude cut-off (and hence more events) and because they evaluate Kagans angles to isolate interplate aftershocks from those occurring within either the subducting or overthrusting crust, both of which show no correlation with the co-seismic rupture area (see figure 3b and c of ref. 65).

These aftershock locations and the distributed slip models shown in Extended Data Fig. 8 suggest that large co-seismic displacements ( $>20$  m) in the Tohoku-oki earthquake did not continue  $>50$  km southeast of the MTL. **c**, Afterslip. Red arrows show one-year postseismic displacements of seafloor GPS sites. Blue arrows show predicted GPS vectors from the viscoelastic model of Sun and Wang<sup>24</sup>. Thick grey contours (numbers are in metres) show the distribution of afterslip<sup>24</sup>. In the dip direction, shallow afterslip is constrained to occur predominantly seaward of site FUKU and thus south of the forearc segment boundary. In the along-strike direction, the northern termination of the afterslip patch is expected to be south of the main rupture area. To the south, the afterslip may extend much farther than depicted by the slip patch shown in Extended Data Fig. 9c and may extend as far south as the Joban seamount chain<sup>24</sup>. The occurrence of rapid afterslip is exactly what would be expected if the region southeast of the forearc segment boundary displayed rate-strengthening behaviour during the Tohoku-oki earthquake.





**Extended Data Figure 10 | Numerical model of earthquake cycles on a fault obeying rate-state friction in the presence of spatially heterogeneous frictional properties.** **a**, Assumed distribution of the rate-dependence parameter  $a - b$ . **b**, Assumed distribution of the coefficient of friction. **c**, Evolution of fault slip in space and time. Black lines denote interseismic fault slip every 5 years, and red dashed lines denote coseismic slip with a time interval of 2 s. To illustrate the effects of spatial variations in the coefficient of friction and the rate-dependence parameter  $a - b$  on the patterns of seismicity, we performed simulations of earthquake cycles of a fault governed by rate-state friction. We assumed a relatively simple case of two asperities (high stress, strong velocity-weakening fault sections) separated by a weak (low coefficient of friction, weak velocity-weakening fault section). The computational domain is 100 km long, and the characteristic size of asperities is 20 km. Simulations were performed using a boundary integral method<sup>66,67</sup>. The fault is driven at a fault distance of 100 km by prescribing a constant velocity of  $100 \text{ mm yr}^{-1}$ . We assumed a constant normal stress of 50 MPa, and slip-weakening

displacement of 10 mm. Unlike the case of a single velocity-weakening asperity that evolves to a sequence of characteristic earthquakes, the modelled earthquake sequence reveals a rich complexity and resembles many features of seismicity in the Tohoku area. There are a number of sub-events of variable size that nucleate predominantly at the boundaries of high-strength asperities, but are arrested before they grow into system-size earthquakes. These sub-events may be analogous to large ( $M_w 7$ ) earthquakes that occurred in the Tohoku area before and after the great 2011 earthquake. Occasionally, the entire area breaks in a mega-event (for example, between 60 m and 70 m of cumulative slip). The slip magnitude in these events is determined by the prior slip history and pre-stress. The model also predicts episodic creep in the middle of the velocity-weakening patch (for example, at 38 m of cumulative slip between 45 km and 60 km) that may be relevant to inferences of low seismic coupling of certain parts of the megathrust. Additional complexity in slip behaviour is likely to be introduced by variations in frictional properties at different spatial wavelengths, and in both along-strike and down-dip directions.

# Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain

Fumiaki Sugahara<sup>1,2</sup>, Juan Pascual-Anaya<sup>1</sup>, Yasuhiro Oisi<sup>3</sup>, Shigehiro Kuraku<sup>4</sup>, Shin-ichi Aota<sup>1</sup>, Noritaka Adachi<sup>5</sup>, Wataru Takagi<sup>1</sup>, Tamami Hirai<sup>1</sup>, Noboru Sato<sup>6</sup>, Yasunori Murakami<sup>7</sup> & Shigeru Kuratani<sup>1</sup>

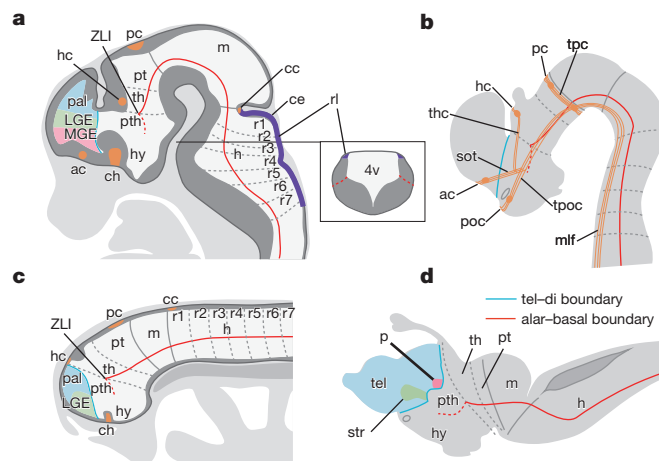
The vertebrate brain is highly complex, but its evolutionary origin remains elusive. Because of the absence of certain developmental domains generally marked by the expression of regulatory genes, the embryonic brain of the lamprey, a jawless vertebrate, had been regarded as representing a less complex, ancestral state of the vertebrate brain. Specifically, the absence of a *Hedgehog*- and *Nkx2.1*-positive domain in the lamprey subpallium was thought to be similar to mouse mutants in which the suppression of *Nkx2-1* leads to a loss of the medial ganglionic eminence<sup>1,2</sup>. Here we show that the brain of the inshore hagfish (*Eptatretus burgeri*), another cyclostome group, develops domains equivalent to the medial ganglionic eminence and rhombic lip, resembling the gnathostome brain. Moreover, further investigation of lamprey larvae revealed that these domains are also present, ruling out the possibility of convergent evolution between hagfish and gnathostomes. Thus, brain regionalization as seen in crown gnathostomes is not an evolutionary innovation of this group, but dates back to the latest vertebrate ancestor before the divergence of cyclostomes and gnathostomes more than 500 million years ago.

During development, the vertebrate brain exhibits highly organized regionalization along the dorsoventral and anteroposterior axes<sup>3,4</sup>; the entire neural tube is divided dorsoventrally into basal and alar plates, medially united by the roof plate dorsally, and by the floor plate ventrally. Anteroposteriorly, the brain primordium is divided into a series of 'neuromeres', the boundaries of which serve as scaffolds for axonal growth, leading to conserved patterns of nerve tracts (Fig. 1a, b and Extended Data Fig. 1). Each domain is determined through specific and regionalized gene expression to differentiate into a functionally specialized region. This brain regionalization, with distinct domains marked by differential, specific gene expression patterns (or genoarchitecture)<sup>4</sup>, is surprisingly conserved in the gnathostome embryos examined so far (Fig. 1a, b), showing that the basic developmental plan of this organ has a deep evolutionary origin and continues to serve as a blueprint for brain development across the gnathostomes. However, it remains unclear when this developmental programme emerged in evolution.

Previous studies on the brain of the lamprey, one of the two groups of cyclostomes together with the hagfish<sup>5</sup>, have suggested that this might represent a pre-gnathostome state of brain evolution, mainly because of the absence of two key domains: (1) the pallidum, which is involved in motor control and develops from a *Nkx2.1*-positive region within the subpallium in gnathostomes, the medial ganglionic eminence (MGE)<sup>4</sup> (Fig. 1a, c); and (2) a morphologically distinct cerebellum, which has crucial roles in sensory integration and motor planning, as well as complex cognitive processing in gnathostomes<sup>6</sup> (Fig. 1a, c). The absence of such structures in the lamprey has placed the origin of the complete vertebrate brain architecture, with an MGE and cerebellum, at the base of the gnathostome lineage<sup>2,7</sup>. However, GABAergic ( $\gamma$ -aminobutyric-acid-releasing) interneurons, which in gnathostomes

migrate from the embryonic MGE and pallidum-like domains, have been observed in juvenile and adult lampreys<sup>8–10</sup> (Fig. 1d), raising the possibility that the embryonic brain of the lamprey represents a secondary reduction from the ancestral one<sup>7</sup>.

To address this question, we studied a member of a different group of cyclostomes, the inshore hagfish (*E. burgeri*). The adult brain of the hagfish has a greatly reduced brain ventricle and exhibits highly differentiated organization of the pallium, which makes it difficult to compare with that of other vertebrates<sup>11</sup> (Extended Data Fig. 2a, b).



**Figure 1 | Regionalization of the vertebrate brains.** **a**, Schematic drawing of the embryonic gnathostome brain based on a mouse embryonic day (E)12.5 embryo. Inset shows a transverse section at the level of the hindbrain showing the position of the rhombic lip. **b**, Nerve tracts and commissures in the gnathostome embryonic brain based on the shark embryo<sup>19</sup> (Extended Data Fig. 1). **c**, Embryonic lamprey brain as revealed by previous studies<sup>2</sup>. **d**, Adult brain of the lamprey, based on refs 10 and 20. Red dotted lines indicate the alar plate-basal plate boundary (sulcus limitans). Blue lines delimit the telencephalon-diencephalon (tel-di) boundary. Orange circles in **b** indicate the position of cell bodies of the early nerve tracts<sup>19</sup>. Coloured regions in **c** correspond to the same colours in **d**. Note that although the pallidum equivalent region ('p') was suggested to be present in the adult lamprey brain, its embryonic primordium (MGE: *Nkx2.1*-positive region) has not been reported so far. ac, anterior commissure; cc, cerebellar commissure; ce, cerebellum; ch, optic chiasm; h, hindbrain; hc, habenular commissure; hy, hypothalamus; m, midbrain; LGE, lateral ganglionic eminence; mlf, medial longitudinal fascicle; p, prospective pallidum; pal, pallium; pc, posterior commissure; poc, postoptic commissure; pt, pretectum; pth, prethalamus; r1–7, rhombomeres 1–7; rl, rhombic lip; sot, supraoptic tract; str, striatum; tel, telencephalon; th, thalamus; thc, tract of the habenular commissure; tpc, tract of the posterior commissure; tpoc, tract of the postoptic commissure; 4v, fourth ventricle.

<sup>1</sup>Evolutionary Morphology Laboratory, RIKEN, Kobe 650-0047, Japan. <sup>2</sup>Division of Biology, Hyogo College of Medicine, Nishinomiya 663-8501, Japan. <sup>3</sup>Development and Function of Inhibitory Neural Circuits, Max Planck Florida Institute for Neuroscience, Jupiter, Florida 33458, USA. <sup>4</sup>Phyloinformatics Unit, RIKEN Center for Life Science Technologies, Kobe 650-0047, Japan.

<sup>5</sup>Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA. <sup>6</sup>Division of Gross Anatomy and Morphogenesis, Niigata University Graduate School of Medical and Dental Sciences, Niigata 950-8510, Japan. <sup>7</sup>Graduate School of Science and Engineering, Ehime University, Matsuyama 790-8577, Japan.

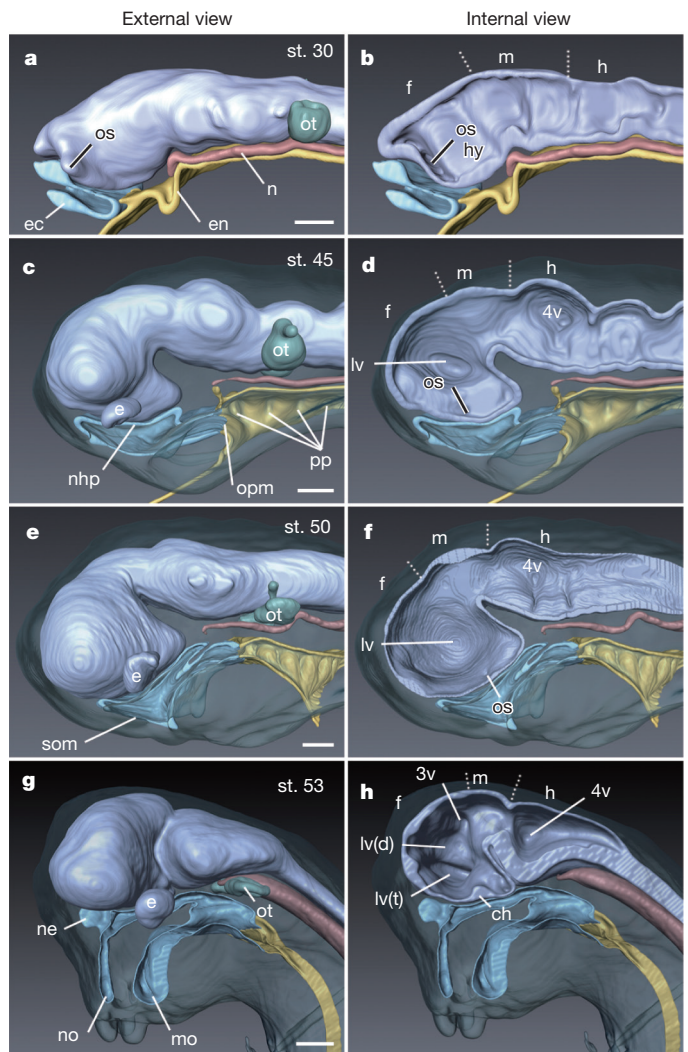


Moreover, as is the case for the lamprey, neither a pallidum nor a cerebellum has been identified in the hagfish<sup>12,13</sup>. The identification of embryonic domains in the hagfish brain has also been difficult, mainly because of the distortion apparent in fixed specimens<sup>14–16</sup>. To overcome these problems, we obtained embryos of *E. burgeri* at Bashford Dean stages 45 and 53 (refs 17 and 18; Extended Data Fig. 2c–h), and compared brain development with lampreys, as well as with the cloudy catshark (*Scyliorhinus torazame*), as a gnathostome outgroup to cyclostomes.

Observations of the developing hagfish brain (Fig. 2 and Extended Data Fig. 3) prompted us to consider putative homologies of hagfish brain regionalization with those of other vertebrates. At stage 53, the hagfish brain clearly shows well-differentiated domains with conspicuous ventricles similar to those in gnathostomes (Fig. 2h), which we describe here according to the prosomeric model<sup>3</sup>. In gnathostomes and the lamprey, transverse commissures serve as landmarks for the identification of neuromeres<sup>11,19,20</sup> (Fig. 1a–c). Longitudinal nerve tracts, such as the medial longitudinal fasciculus and some commissures—including the optic chiasm and the habenular and posterior commissures—allowed us to mark the regionalization of the hagfish embryonic brain. In this, the habenular commissure defines the roof of the thalamus, or prosomere 2 (p2)<sup>3</sup>, and the posterior commissure marks the posterior boundary of the pretectum (p1)<sup>4,11</sup> (Figs 1 and 3a, h). The ventral limit of the telencephalon is defined as dorsal to the optic chiasm (Fig. 3a). The interbulbar commissure, already described in the lamprey<sup>11</sup>, is also present in the hagfish (Fig. 3a, h). The latter commissure is equivalent to the gnathostome pallial or hippocampal commissures in the telencephalon, and crosses secondary olfactory fibres as in the lamprey<sup>11</sup>. Finally, the commissure lying posterior to the midbrain–hindbrain boundary of the hagfish brain corresponds to the commissura vestibulo-lateralis, which consists of the eighth nerve and lateral line nerve fibres as in the lamprey, although in the hagfish the cerebellar commissure has not been identified<sup>6</sup> (Fig. 3a). Throughout the developmental stages examined, we could not identify any overt epiphysis in the hagfish, although this has been described in the lamprey and jawed vertebrates (Extended Data Fig. 4a–g).

To test for the presence or absence of neuromeres, we investigated the brain genoarchitecture of the hagfish embryo to evaluate its regionalization. The telencephalon of the hagfish embryo could be identified as a domain expressing *FoxG1*, *EmxB* and *Pax6* orthologues<sup>21,22</sup> (Fig. 3c–e, j–l and Extended Data Figs 2k–m, r–t and 5). The hypothalamus was located rostral to the diencephalon, identified by the co-expression of one hagfish *Hedgehog* gene, *Hh2* (one of the three hagfish *Hh* genes newly identified in this study, named *Hh2–4*, and distinct from our previously reported hagfish *Hh1* (ref. 18); see Methods and Extended Data Figs 5 and 6), and *Nkx2.1/2.4*, an orthologous gene of both gnathostome *Nkx2.1* and *Nkx2.4* genes (Fig. 3f, g and Extended Data Figs 2n, o and 5). However, we could not determine whether the anterior end of the floor plate was hypothalamic or not<sup>23</sup>. As in the lamprey<sup>2</sup>, hagfish *Hh2* expression also clearly marked the position of the zona limitans intrathalamica (ZLI) between the prethalamus (p3) and thalamus (p2). This position had not been identified previously (Extended Data Fig. 4h–k). Given the *Shh*-dependent inductive activity of the ZLI<sup>24</sup>, its presence is consistent with the p2–p3 boundary in the hagfish diencephalon as noted earlier. More caudally, the pretectum (p1), which is identified by the posterior commissure, expressed *Pax6*, as in gnathostomes<sup>4</sup> (Fig. 3a, k and Extended Data Fig. 2i, t). Thus, in the hagfish, we clearly identified the hypothalamus and diencephalic prosomeres (pretectum (p1), thalamus (p2) and prethalamus (p3)) rostral to the midbrain, with the ZLI between p2 and p3. Finally, the mesencephalon was defined as a *Pax6*-negative region caudal to the posterior commissure (Fig. 3a, h, k and Extended Data Fig. 2i, t).

Notably, and unlike in lampreys<sup>25</sup>, we detected in the stage 53 *E. burgeri* embryo an independent and conspicuous expression domain of *Nkx2.1/2.4*, dorsal (rostral according to conventional columnar model<sup>4</sup>) to the optic chiasm, corresponding to the region where the



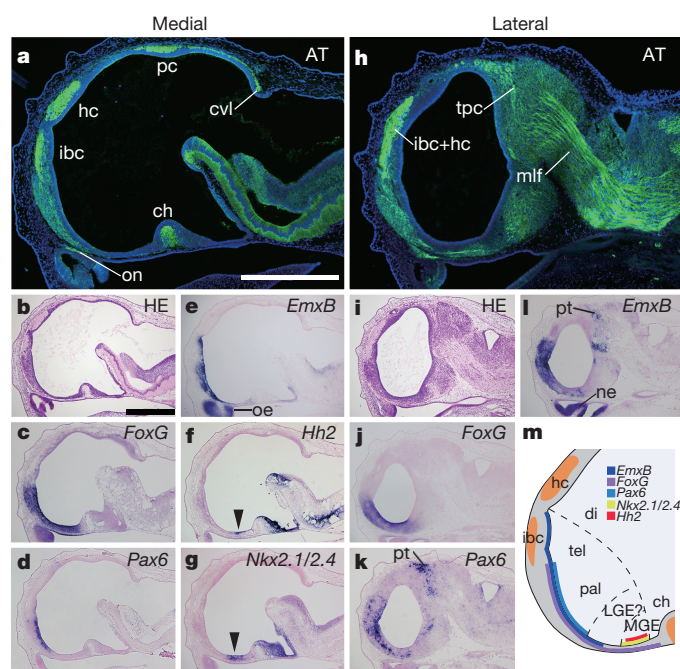
**Figure 2 | Development of the ventricular system in the hagfish brain.**

**a–h**, Three-dimensional images reconstructed from serial sections of *E. burgeri* embryos with Avizo software. External (**a**, **c**, **e** and **g**) and internal (**b**, **d**, **f** and **h**) views are shown. The central nervous systems are in purple; ectoderm is in light blue and green; endoderm is in yellow; and notochord is in orange. The developmental stages are based on ref. 18. Identification of brain regions is based on brain morphology, nerve staining and comparative gene expression patterns (Fig. 3 and Extended Data Figs 2 and 3). At stage (st.) 53, lateral ventricles can be divided into a telencephalic and a diencephalic part (lv(t) and lv(d), respectively)<sup>6</sup>. e, eye; ec, ectoderm; en, endoderm; f, forebrain; lv, lateral ventricle; mo, mouth opening; n, notochord; ne, nasal epithelium; nhp, naso-hypophyseal plate; no, nasal opening; os, optic stalk; ot, otic vesicle; opm, oropharyngeal membrane; pp, pharyngeal pouches; som, secondary oral membrane; 3v, third ventricle. See Fig. 1 for other abbreviations. Scale bars, 200  $\mu$ m.

gnathostome MGE would form<sup>22</sup> (Fig. 3g). Moreover, *Hh2* was also expressed in a slightly restricted area of this *Nkx2.1/2.4*-positive domain, possibly corresponding to a part of the MGE and preoptic area<sup>4</sup> (Fig. 3f). Therefore, the putative MGE in the hagfish arises as a conspicuous region within the brain, located in the rostral part of the telencephalon (ventral according to the conventional columnar model<sup>4</sup>), and possibly associated with the lateral ganglionic eminence posterior to it (Fig. 3m).

In modern gnathostomes, the cerebellum differentiates from the dorsal part of rhombomere 1, whereby granular cells are supplied from the rhombic lip to constitute a layer of the cerebellar cortex<sup>4</sup>. The rhombic lip in gnathostomes expresses *Pax6*, which enables granular cells to migrate<sup>26</sup>. Therefore, regional specification of the rhombic lip is a

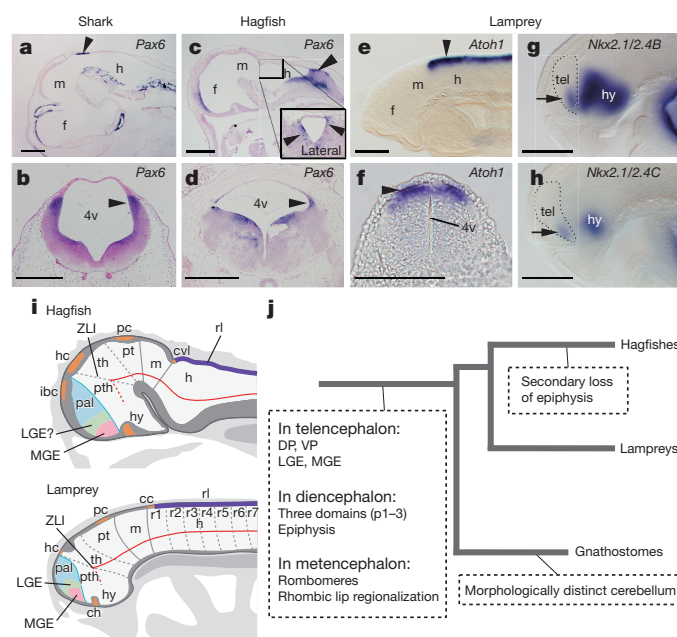




**Figure 3 | Gene expression and nervous staining in *E. burgeri* at stage 53.** **a, h**, Immunohistochemical staining of the axon bundles by anti-acetylated tubulin antibody (green) and DAPI (blue). **b, i**, Haematoxylin and eosin (HE) staining. **c–g, j–l**, Gene expression patterns involved in the regionalization of the forebrain. Arrowheads in **f** and **g** indicate expressions in the MGE region of the telencephalon. **m**, Schematic drawing of gene expressions in the telencephalon. The telencephalic territory can be identified from the expressions of *FoxG*, *Pax6* and *EmxB*<sup>21,22</sup>. cvl, commissura vestibulo-lateralis; di, diencephalon; ibc, interbulbar commissure; on, olfactory nerve. See Fig. 1 for other abbreviations. Scale bars, 500  $\mu$ m.

prerequisite to acquisition of the cerebellum. However, an overt cerebellum is missing in cyclostomes, although there are fibres possibly equivalent to commissures posterior to the midbrain–hindbrain boundary in gnathostomes<sup>6</sup> (Fig. 1). In the lamprey, previous studies have deduced the absence of a rhombic-lip-derived cerebellum from the lack of *Pax6* expression in the corresponding area<sup>2</sup>. Unlike in the lamprey, the hagfish rhombencephalon clearly develops a *Pax6*-positive rhombic lip, as in gnathostomes (Fig. 4a–d). This was confirmed by the expression of a different rhombic lip marker, *Atoh1*, which is involved in the development of excitatory cells in the gnathostome cerebellum, in the hagfish dorsal hindbrain<sup>27</sup> (Extended Data Fig. 7n).

Our results show that the hagfish and gnathostome embryonic brains share a common regional patterning, leading us to two scenarios: (1) the hagfish/gnathostome genoarchitecture represents an ancestral programme for all vertebrates, and lampreys have lost the MGE and rhombic lip secondarily; or (2) the ancestral brain developmental pattern did not have an MGE or rhombic lip, which were acquired independently in the lineages of hagfish and gnathostomes by convergent evolution. To clarify this, we have reinvestigated brain development of the Japanese lamprey (*Lethenteron japonicum*) with a recently reported draft genome sequence<sup>28</sup>. We found two extra lamprey *Nkx2.1* orthologous genes (*Nkx2.1/2.4B* and *Nkx2.1/2.4C*; see Methods, Extended Data Figs 5 and 8 and Extended Data Table 1), as well as one extra *Hedgehog* paralogue, *HhD*, in addition to the previously reported *Nkx2.1* (renamed as *Nkx2.1/2.4A*) and *HhA–C* genes<sup>7</sup> (Extended Data Figs 5 and 9). Surprisingly, lamprey *Nkx2.1/2.4B* and *Nkx2.1/2.4C* were expressed in a rostral domain within the subpallium, suggesting the presence of an MGE in the lamprey (Fig. 4g, h and Extended Data Fig. 9). On the other hand, no *HhD* expression was detected (Extended Data Fig. 9), suggesting that, although present, specification of the MGE in the lamprey is still somehow divergent compared with other vertebrates.



**Figure 4 | Presence of the rhombic lip and MGE in the hagfish and lamprey brains, and an evolutionary model of the vertebrate brain.**

**a–f**, *Pax6* expression in the hindbrain of a stage 31 catshark embryo (**a, b**) and a hagfish stage 53 embryo (**c, d**). *Atoh1* expression in a lamprey stage 26 embryo (**e, f**). Sagittal views (**a, c, e**) and transverse sections (**b, d, f**) are shown. Arrowheads indicate expressions in the rhombic lip. Inset in **c** shows rhombic lip expression on a lateral section. **g, h**, *Nkx2.1/2.4* homologue gene expression patterns in a lamprey stage 27 embryo. Arrows indicate telencephalic expression of *Nkx2.1/2.4B* and *C*. **i**, Schematic drawing of embryonic hagfish and lamprey brains (stages 53 and 26, respectively) as revealed by the present study. **j**, An evolutionary model of the vertebrate brain. DP, dorsal pallium; fp, floor plate; VP, ventral pallium. See Figs 1–3 for other abbreviations. Scale bars, 500  $\mu$ m (**a–d**) and 100  $\mu$ m (**e–h**).

Next, we looked for the presence of a rhombic lip in *L. japonicum* larvae. Lamprey *Atoh1*, *Wnt1* and *Ptf1a* genes, as well as the recently identified lamprey *Pax6B* gene<sup>29</sup>, were expressed in a dorsoventrally organized pattern in the hindbrain (Fig. 4), indicating specification of a rhombic-lip-like property as in gnathostomes, although *Pax6B* expression was found more ventrally than in gnathostomes (Extended Data Fig. 7s). Therefore, as in the case of the MGE, lamprey rhombic lip specification is divergent from the gnathostome condition. Notably, because the gnathostome *Atoh1* and *Ptf1a* genes are involved in the development of the cerebellum granule and Purkinje cells, respectively<sup>27</sup>, certain genetic backgrounds underlying the acquisition of the cerebellum proper in gnathostomes, yet absent in cyclostomes, might have been established before the splitting of the two groups (Fig. 4). Altogether, the lamprey developmental pattern suggests the presence of a rhombic lip domain (Fig. 4 and Extended Data Fig. 7).

The absence of certain domains in the lamprey brain once led researchers to examine a model in which the brain evolution followed a stepwise elaboration towards crown gnathostomes. However, we show here that two major domains of the vertebrate brain, previously regarded as gnathostome novelties, are indeed present in cyclostomes as distinct gene expression domains. These domains would have followed further subsequent elaboration in each lineage. For instance, only the gnathostome rhombic lip led to the cerebellum proper as a gnathostome novelty (Fig. 4j). That part would probably have preceded acquisition of a jaw, because jawless stem gnathostomes, such as ostracoderms, probably had a morphologically distinct cerebellum<sup>30</sup>. Thus, our findings markedly alter the accepted evolutionary model (Fig. 4) and suggest that the vertebrate brain genoarchitecture, with regional gene expression patterns as seen in crown gnathostomes, is much

older than previously thought, dating back to more than 500 million years ago when the latest common ancestor of crown vertebrates arose (Fig. 4).

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 13 April; accepted 7 December 2015.**

**Published online 15 February 2016.**

- Sussel, L., Marin, O., Kimura, S. & Rubenstein, J. L. R. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development* **126**, 3359–3370 (1999).
- Murakami, Y., Uchida, K., Rijli, F. M. & Kuratani, S. Evolution of the brain developmental plan: insights from agnathans. *Dev. Biol.* **280**, 249–259 (2005).
- Rubenstein, J. L. R., Martinez, S., Shimamura, K. & Puelles, L. The embryonic vertebrate forebrain: the prosomeric model. *Science* **266**, 578–580 (1994).
- Watson, C., Paxinos, G. & Puelles, L. *The Mouse Nervous System* (Academic Press, 2011).
- Kuraku, S. Insights into cyclostome phylogenomics: pre-2R or post-2R? *Zool. Sci.* **25**, 960–968 (2008).
- Nieuwenhuys, R., ten Donkelaar, H. J. & Nicholson, C. *The Central Nervous System of Vertebrates* (Springer, 1998).
- Sugahara, F., Murakami, Y., Adachi, N. & Kuratani, S. Evolution of the regionalization and patterning of the vertebrate telencephalon: what can we learn from cyclostomes? *Curr. Opin. Genet. Dev.* **23**, 475–483 (2013).
- Marin, O. & Rubenstein, J. L. R. A long, remarkable journey: tangential migration in the telencephalon. *Nature Rev. Neurosci.* **2**, 780–790 (2001).
- Pombal, M. A., Alvarez-Otero, R., Perez-Fernandez, J., Solveira, C. & Megias, M. Development and organization of the lamprey telencephalon with special reference to the GABAergic system. *Front. Neuroanat.* **5**, 20 (2011).
- Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B. & Grillner, S. Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Curr. Biol.* **21**, 1081–1091 (2011).
- Wicht, H. & Northcutt, R. G. The forebrain of the Pacific hagfish: a cladistic reconstruction of the ancestral craniate forebrain. *Brain Behav. Evol.* **40**, 25–64 (1992).
- Wicht, H. & Northcutt, R. G. An immunohistochemical study of the telencephalon and the diencephalon in a myxinoideous jawless fish, the Pacific hagfish, *Eptatretus stouti*. *Brain Behav. Evol.* **43**, 140–161 (1994).
- Larsell, O. *The Comparative Anatomy and Histology of the Cerebellum*. (Univ. Minnesota Press 1967).
- Conel, J. L. The development of the brain of *Bdellostoma stouti* 1. External growth changes. *J. Comp. Neurol.* **47**, 343–403 (1929).
- von Kupffer, C. *Studien zur vergleichenden Entwicklungsgeschichte des Kopfes der Cranioten, Heft 4: Zur Kopfentwicklung von Bdellostoma* (Lehmann, 1900).
- Ota, K. G., Kuraku, S. & Kuratani, S. Hagfish embryology with reference to the evolution of the neural crest. *Nature* **446**, 672–675 (2007).
- Dean, B. *On the Embryology of Bdellostoma stouti: a General Account of Myxinoideous Development from the Egg and Segmentation to Hatching* 220–276 (Harvard Univ., 1899).
- Oisi, Y., Ota, K. G., Kuraku, S., Fujimoto, S. & Kuratani, S. Craniofacial development of hagfishes and the evolution of vertebrates. *Nature* **493**, 175–180 (2013).
- Barreiro-Iglesias, A., Villar-Cheda, B., Abalo, X. M., Anadon, R. & Rodicio, M. C. The early scaffold of axon tracts in the brain of a primitive vertebrate, the sea lamprey. *Brain Res. Bull.* **75**, 42–52 (2008).
- Pombal, M. A., Megias, M., Bardet, S. M. & Puelles, L. New and old thoughts on the segmental organization of the forebrain in lampreys. *Brain Behav. Evol.* **74**, 7–19 (2009).
- Tank, E. M. et al. Patterns and consequences of vertebrate *Emx* gene duplications. *Evol. Dev.* **11**, 343–353 (2009).
- Hébert, J. M. & Fishell, G. The genetics of early telencephalon patterning: some assembly required. *Nature Rev. Neurosci.* **9**, 678–685 (2008).
- Puelles, L. & Rubenstein, J. L. R. A new scenario of hypothalamic organization: rationale of new hypotheses introduced in the updated prosomeric model. *Front. Neuroanat.* **9**, 27 (2015).
- Kiecker, C. & Lumsden, A. The role of organizers in patterning the nervous system. *Annu. Rev. Neurosci.* **35**, 347–367 (2012).
- Sugahara, F. et al. Involvement of Hedgehog and FGF signalling in the lamprey telencephalon: evolution of regionalization and dorsoventral patterning of the vertebrate forebrain. *Development* **138**, 1217–1226 (2011).
- Engelkamp, D., Rashbass, P., Seawright, A. & van Heyningen, V. Role of Pax6 in development of the cerebellar system. *Development* **126**, 3585–3596 (1999).
- Wullmann, M. F. et al. The long adventurous journey of rhombic lip cells in jawed vertebrates: a comparative developmental analysis. *Front. Neuroanat.* **5**, 27 (2011).
- Mehta, T. K. et al. Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). *Proc. Natl Acad. Sci. USA* **110**, 16044–16049 (2013).
- Feiner, N., Meyer, A. & Kuraku, S. Evolution of the vertebrate Pax4/6 class of genes with focus on its novel member, the Pax10 gene. *Genome Biol. Evol.* **6**, 1635–1651 (2014).
- Stensiö, E. A. *The Downtonian and Devonian Vertebrates of Spitsbergen: Part 1, Family Cephalaspidae* (I kommisjon hos J. Dybwad, 1927).

**Acknowledgements** We thank O. Kakitani for hagfish sampling; K. Shirato for shark sampling; S. Shibuya, K. Yamamoto and Y. Yamamoto for maintenance of aquarium tanks; D. Sipp for critical reading of this manuscript; C. Mitgutsch and M. Kawaguchi for discussions; K. G. Ota and I. Sato for technical support; and H. Nagashima for lamprey sampling and discussions. This research was supported by direct budget supplied by Centre for Developmental Biology, RIKEN, JSPS KAKENHI grant numbers 15H02416 and 25840133, and by Hyogo Science and Technology Association.

**Author Contributions** F.S., J.P.-A., Y.M. and S. Kuratani designed the experiments and wrote the paper. F.S., J.P.-A., Y.O., S.A., N.A. and T.H. performed molecular works. N.S. discussed and interpreted the data. S. Kuraku performed the molecular phylogenetic analysis. W.T. performed the *Nkx2.1/2.4* locus synteny conservation analysis.

**Author Information** The sequences of the genes described in this study have been deposited in the NCBI GenBank under accessions LC028239–LC028243, LC028245 and KT897926–KT897927 and KT897930–KT897938. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S. Kuratani (saizo@cdb.riken.jp).



## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

### Sample collection and aquarium maintenance of hagfish, lamprey and shark.

Mature males and females of the Japanese inshore hagfish (*E. burgeri*), were obtained in the Japan Sea off Shimane prefecture as described<sup>31</sup>. Eggs were deposited in a cage set on the sea bottom in October 2011. Deposited eggs were then collected and incubated in an aquarium until embryos reached the required stages. Embryos were fixed with 4% paraformaldehyde in late March 2012. We mainly used two embryos (at stages 45 and 53) in this study. Embryos used in Fig. 4d and Extended Data Fig. 7m, n were obtained previously<sup>16,18</sup>. Sources of the Japanese lamprey (*L. japonicum*; recently included as a synonym for *L. camtschaticum*), were as described<sup>32</sup>. Lamprey embryos were staged as described<sup>33</sup> and fixed in 4% paraformaldehyde. Adult catsharks (*S. torazame*), were obtained in Ibaraki prefecture, Japan, and kept in aquarium tanks. Deposited eggs were incubated in the same tanks at 16 °C until embryos reached the required stages. Catshark embryos were staged and fixed with 4% paraformaldehyde as described<sup>34</sup>. The sampling and experiment were conducted according to the institutional and national guidelines for animal ethics.

**Gene identification, cDNA cloning and sequencing.** For the identification of hagfish *Hedgehog*, *Atoh1* and catshark *Atoh1*, *Ptf1a* and *Wnt1* genes, the nucleotide sequences were obtained from embryonic transcriptomes of *E. burgeri* and *S. torazame*, respectively, and assembled from an in-house RNA-sequencing (RNA-seq) data set (to be published elsewhere). Hagfish *FoxG* sequence was retrieved from the Vertebrate Time Capsule database<sup>35</sup>. In all cases, we performed TBLASTN using as queries the amino acid sequences of corresponding orthologues from human, mouse, chicken and zebrafish (plus *L. japonicum* *HhA*, *-B*, *-C* and the amphioxus *Branchiostoma floridae* *Hh* gene in the case of hagfish *Hedgehog*). We found four putative hagfish *Hh* genes, including the previously reported *Hh1* (ref. 18), and designated the three new genes as *Hh2–4* arbitrarily. The Lamprey *Nkx2.1/2.4*, *Atoh1*, *Ptf1a* and *Wnt1* genes were identified by means of TBLASTN using human, mouse, chicken and zebrafish counterparts as queries against *L. japonicum* low-coverage draft genome sequence version LetJap1.0 (ref. 28; accession number APJL000000000, <http://l.jlampreygenome.imcb.a-star.edu.sg/>). We found three *Nkx2.1* or *Nkx2.4* gene-containing scaffolds: 242 (GenBank accession number KE993913, containing the previously reported TTF-1; ref. 36), 17 (KE993688) and 73 (KE993744). We found lamprey *Wnt1* in scaffold 103 (KE993774), two *Ptf1a* paralogues in scaffolds 196 (*Ptf1a-A*; KE993867) and 473 (*Ptf1a-B*; KE994144), and *Atoh1* in scaffold 280 (KE993951). The corresponding gene sequences were predicted by means of GeneWise<sup>37</sup> (<http://www.ebi.ac.uk/Tools/psa/genewise/>) or GENSCAN<sup>38</sup> (<http://genes.mit.edu/GENSCAN.html>) and curated manually. Phylogenetic and synteny conservation analyses (see below) could not resolve the orthology relationships between *Nkx2.1* and *Nkx2.4*, so we designated these genes as *Nkx2.1/2.4A* (TTF-1, scaffold 242), *Nkx2.1/2.4B* (scaffold 17) and *Nkx2.1/2.4C* (scaffold 73). As in gnathostomes<sup>39</sup>, *Nkx2.1/Nkx2.4A* and *C* were linked to the *Nkx2.2* and *Nkx2.8* genes, which were designated as *Nkx2.2/2.8A* and *C*, respectively. A partial sequence of a third putative *Nkx2.2/2.8B* gene was found in contig 79345 (accession number APJL01110193). The shark *FoxG1* homologous gene was cloned using degenerate primers designed based on the consensus amino acid sequences of vertebrate *FoxG1* counterparts (forward primer, 5'-ATGGGNGANMGNAARGA-3'; reverse, 5'-NCCYTGRTTYTGRTGNGT-3'; sequence based on IUPAC ambiguity code). An *E. burgeri* *OtxC* gene fragment was obtained using specific primers based on the nucleotide sequence of the Atlantic hagfish (*Myxine glutinosa*) *OtxC* gene<sup>40</sup>.

Total RNAs of *E. burgeri* and *S. torazame* were extracted from whole embryos using TRIzol reagent (Invitrogen Life Sciences). Reverse transcription PCR (RT-PCR) was performed to amplify fragments of each gene with specific primers. Lamprey gene fragments were obtained from a stage-24 cDNA library (ZAP-cDNA Synthesis Kit, 200400, Stratagene). PCR products were cloned into the pCRII-TOPO vector (Invitrogen) and subsequently sequenced with a 3130 Sequence Analyzer (Applied Biosystems). The 5' and 3' ends were amplified with GeneRacer kits (Invitrogen Life Sciences), or from the lamprey cDNA library with customized primers. The cDNA sequences identified here have been deposited in GenBank under accession numbers LC028239–43, LC028245 and KT897926–27 and KT897930–38. Shark *Pax6* (AB773851), *Dlx2* (AB293582), *Shh* (AB247650), *Nkx2.1* (AB773852), hagfish *Pax6* (AB270704), *EmxA* (AB935431), *EmxB* (AB935432), *Nkx2.1* (AB747372), *Hh1* (AB703680) and *Fgf8/17* (AB703681) and lamprey *Pax6* (AB061220), *Pax6B* (supplementary table 3 of ref. 29), *DlxA* (AB292628), *HhA* (AB124584), *HhB* (AB583549) and *HhC* (AB583548) were obtained previously<sup>7,18,25,41–43</sup>.

**In situ hybridization.** *In situ* hybridization on paraffin wax-embedded sections was performed as described<sup>32</sup>. In brief, deparaffinized sections (8–10 µm) were treated with Protease K (2–10 µg ml<sup>-1</sup>) at room temperature for 5–15 min. Slides were incubated in hybridization buffer (50% formamide, 5 × SSC (pH 7.0), 1% SDS, 50 µg ml<sup>-1</sup> total yeast RNA, 50 µg ml<sup>-1</sup> heparin sulphate, 5 mM EDTA (pH 8.0), 0.1% CHAPS and 1% blocking reagent (Roche)) with 0.2–1.0 µg ml<sup>-1</sup> DIG-labelled RNA probe overnight at 65 °C. For immunohistochemical detection, sections were incubated with alkaline phosphatase (AP)-conjugated anti-DIG Fab fragments (diluted 1:4,000) for 2 h after blocking. Colour development was performed using the BM purple AP substrate (Roche). Whole-mount *in situ* hybridization of lamprey and shark embryos was performed according to refs 32 and 41, respectively.

**Immunohistochemical staining.** Deparaffinized sections of hagfish embryos were blocked with 5% skim milk in TBST (TSTM). They were then reacted overnight at room temperature with an anti-acetylated tubulin antibody (Sigma-Aldrich, T6793) diluted in TSTM (1:1,000). After the samples had been washed with TBST, they were reacted for 2 h with Alexa Fluor 488-conjugated goat anti-mouse antibody (Invitrogen Life Sciences, A11001) diluted in TSTM (1:400). After the final wash in TBST, embryos were observed using an Axio Zoom V16 fluorescence microscope (Carl Zeiss).

**3D reconstruction of the hagfish brain.** The 3D images of the ventricular system of hagfish embryos were reconstructed based on the data set used previously<sup>18</sup>. Reconstructed images were acquired using Avizo software (Visualization Sciences Group).

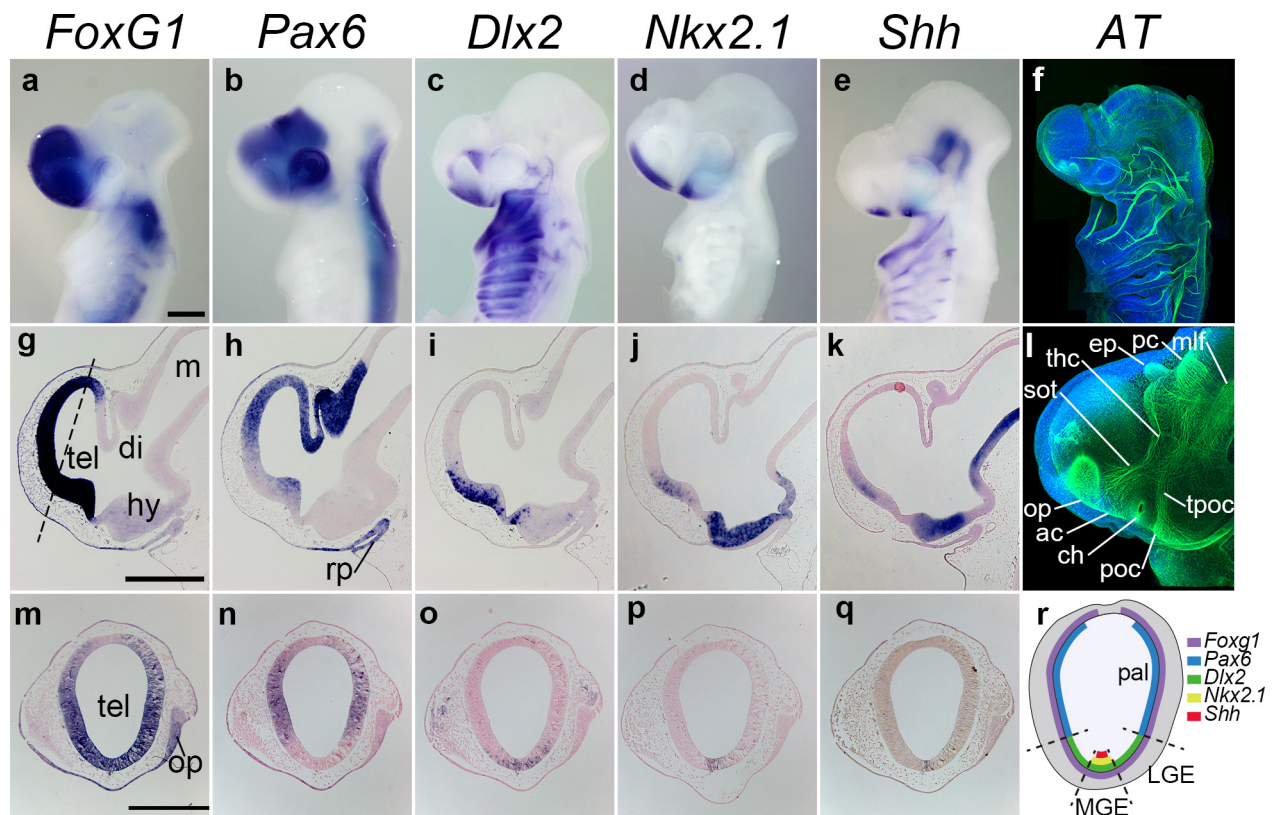
**Molecular phylogenetic analysis.** The molecular phylogenetic trees of *Hh*, *Nkx2.1/2.4*, *Fgf8/17*, *FoxG*, *Wnt1*, *Atoh1* and *Ptf1a* genes of *E. burgeri*, *L. japonicum* and *S. torazame* were inferred with the maximum-likelihood method using PhyML3.0 (<http://www.atgc-montpellier.fr/phyml/download.php>) with the JTT+G4 model.

**Synten conservation analysis.** According to ref. 44, comparisons of genomic regions including vertebrate *Nkx2.1/2.4* were performed among the following species: human (*Homo sapiens*), chicken (*Gallus gallus*), coelacanth (*Latimeria chalumnae*), western clawed frog (*Xenopus tropicalis*), elephant shark (*Callorhynchus milii*) and Japanese lamprey (*L. japonicum*). In brief, gnathostome *Nkx2.1/2.4* orthologues and the adjacent protein-coding genes were arranged using the UCSC (<http://genome.ucsc.edu>) and Ensembl (<http://www.ensembl.org/>) genome browsers. For lamprey, genome scaffolds 17 (GenBank accession KE993688), 73 (KE993744) and 242 (KE993913) containing the *Nkx2.1/2.4* homologues (see above) were analysed using the gene prediction tool, GENESCAN<sup>38</sup>. Subsequently, deduced amino acid sequences of the predicted genes were used as queries for NCBI TBLASTN searches and annotated with an *E*-value cut-off of 1 × 10<sup>-20</sup>.

- Oisi, Y., Kakitani, O., Kuratani, S. & Ota, K. G. in *In Situ Hybridization Methods* (ed. Hauptmann, G.), Ch. 12, 249–262 (Springer, 2015).
- Sugahara, F., Murakami, Y. & Kuratani, S. Gene expression analysis of lamprey embryos. In: *In Situ Hybridization Methods* (ed. Hauptmann, G.), 263–278 (Springer, 2015).
- Tahara, Y. Normal stages of development in the lamprey, *Lampetra reissneri* (Dybowski). *Zoolog. Sci.* **5**, 109–118 (1988).
- Ballard, W. W., Mellinger, J. & Lechenault, H. A series of normal stages for development of *Scyliorhinus canicula*, the lesser spotted dogfish (Chondrichthyes, Scyliorhinidae). *J. Exp. Zool.* **267**, 318–336 (1993).
- Takechi, M. et al. Overview of the transcriptome profiles identified in hagfish, shark, and bichir: current issues arising from some nonmodel vertebrate taxa. *J. Exp. Zool. B Mol. Dev. Evol.* **316**, 526–546 (2011).
- Ogasawara, M., Shigetani, Y., Suzuki, S., Kuratani, S. & Satoh, N. Expression of thyroid transcription factor-1 (TTF-1) gene in the ventral forebrain and endostyle of the agnathan vertebrate, *Lampetra japonica*. *Genesis* **30**, 51–58 (2001).
- Birney, E. & Durbin, R. Using GeneWise in the *Drosophila* annotation experiment. *Genome Res.* **10**, 547–548 (2000).
- Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
- Santagati, F. et al. Identification of *Cis*-regulatory elements in the mouse *Pax9/Nkx2-9* genomic region: implication for evolutionary conserved synteny. *Genetics* **165**, 235–242 (2003).
- Germot, A. et al. Structural evolution of *Otx* genes in craniates. *Mol. Biol. Evol.* **18**, 1668–1678 (2001).
- Adachi, N., Takechi, M., Hirai, T. & Kuratani, S. Development of the head and trunk mesoderm in the dogfish, *Scyliorhinus torazame*: II. Comparison of gene expression between the head mesoderm and somites with reference to the origin of the vertebrate head. *Evol. Dev.* **14**, 257–276 (2012).
- Takechi, M., Adachi, N., Hirai, T., Kuratani, S. & Kuraku, S. The *Dlx* genes as clues to vertebrate genomics and craniofacial evolution. *Semin. Cell Dev. Biol.* **24**, 110–118 (2013).
- Noro, M., Sugahara, F. & Kuraku, S. Reevaluating *Emx* gene phylogeny: homopolymeric amino acid tracts as a potential factor obscuring orthology signals in cyclostome genes. *BMC Evol. Biol.* **15**, 78 (2015).

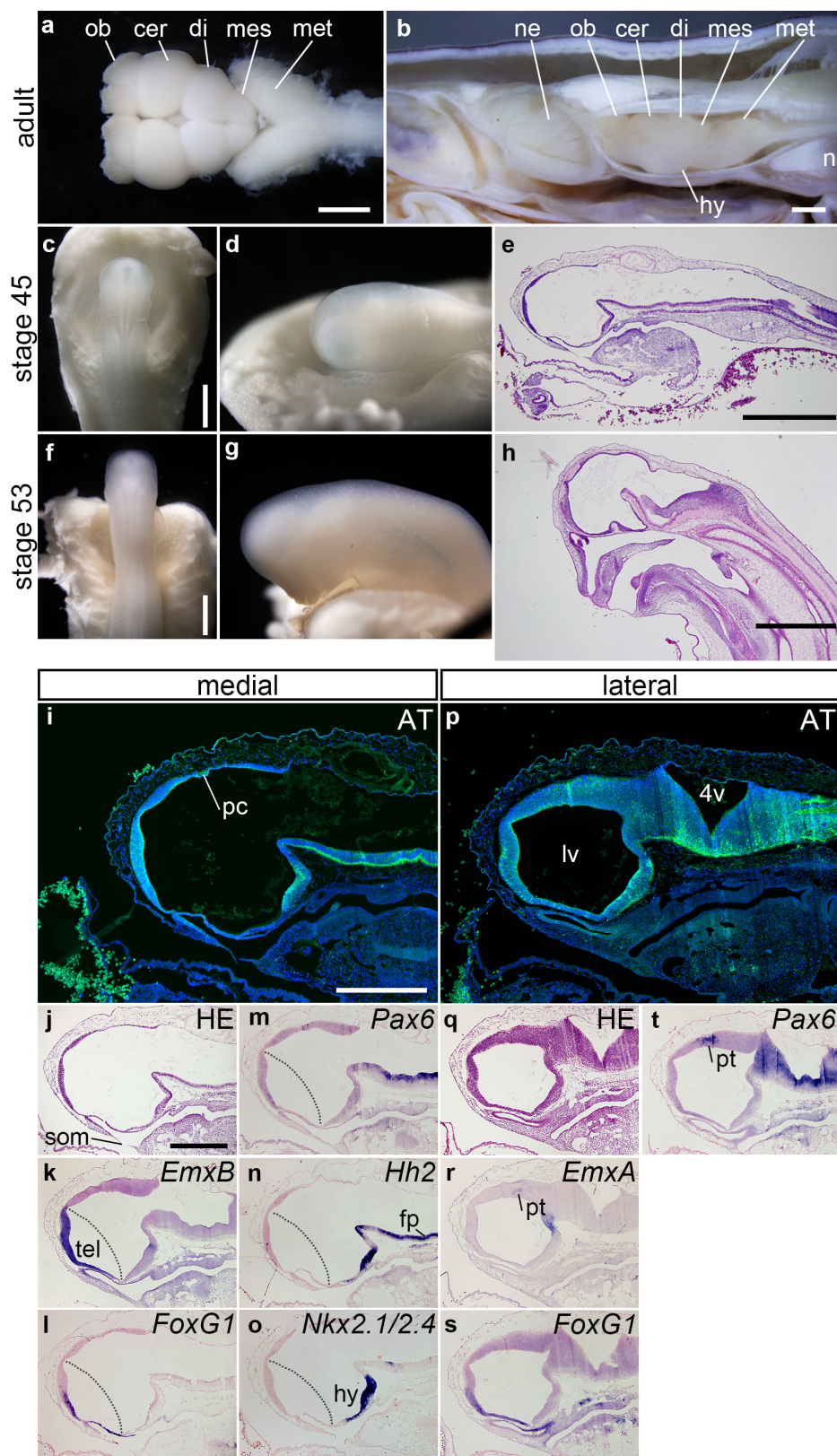


44. Yamaguchi, Y. *et al.* The fifth neurohypophysial hormone receptor is structurally related to the V2-type receptor but functionally similar to V1-type receptors. *Gen. Comp. Endocrinol.* **178**, 519–528 (2012).
45. von Kupffer, K. in *Handbuch der Vergleichenden und Experimentellen Entwicklungslehre der Wirbeltiere* Vol. 2 (ed. Hertwig, O.) 1–272 (G. Fischer, 1906).
46. Conel, J. L. The development of the brain of *Bdellostoma stouti*. II. Internal growth changes. *J. Comp. Neurol.* **52**, 365–499 (1931).
47. Manoli, M. & Driever, W. *nkx2.1* and *nkx2.4* genes function partially redundant during development of the zebrafish hypothalamus, preoptic region, and pallidum. *Front. Neuroanat.* **8**, 145 (2014).



**Extended Data Figure 1 | Embryonic brain of a shark.** a–q, Gene expression patterns and nerve tracts in a stage-27 (ref. 34) embryo of the cloudy catshark (*S. torazame*), to show the conserved developmental pattern of the jawed vertebrate brain. Whole-mount views (a–e), sagittal sections of the brain (g–k) and transverse sections at the telencephalon level (m–q; the sectioned region is shown by the dotted line in g), stained using probes for *FoxG1* (a, g, m), *Pax6* (b, h, n), *Dlx2* (c, i, o),

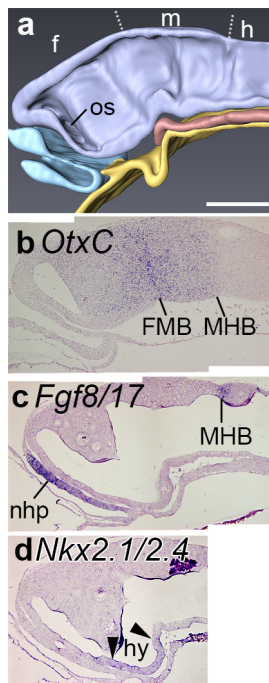
*Nkx2.1* (d, j, p) and *Shh* (e, k, q). f, l, Confocal images of the central and peripheral nervous systems visualized by immunofluorescence using an anti-acetylated tubulin antibody (green) and DAPI (blue). r, Schematic drawing of the shark telencephalon, showing conserved gene expression patterns among jawed vertebrates. Some of these patterns were described previously<sup>7</sup>. op, olfactory placode; rp, Rathke's pouch. See Figs 1 and 2 for other abbreviations. Scale bars, 500  $\mu$ m.



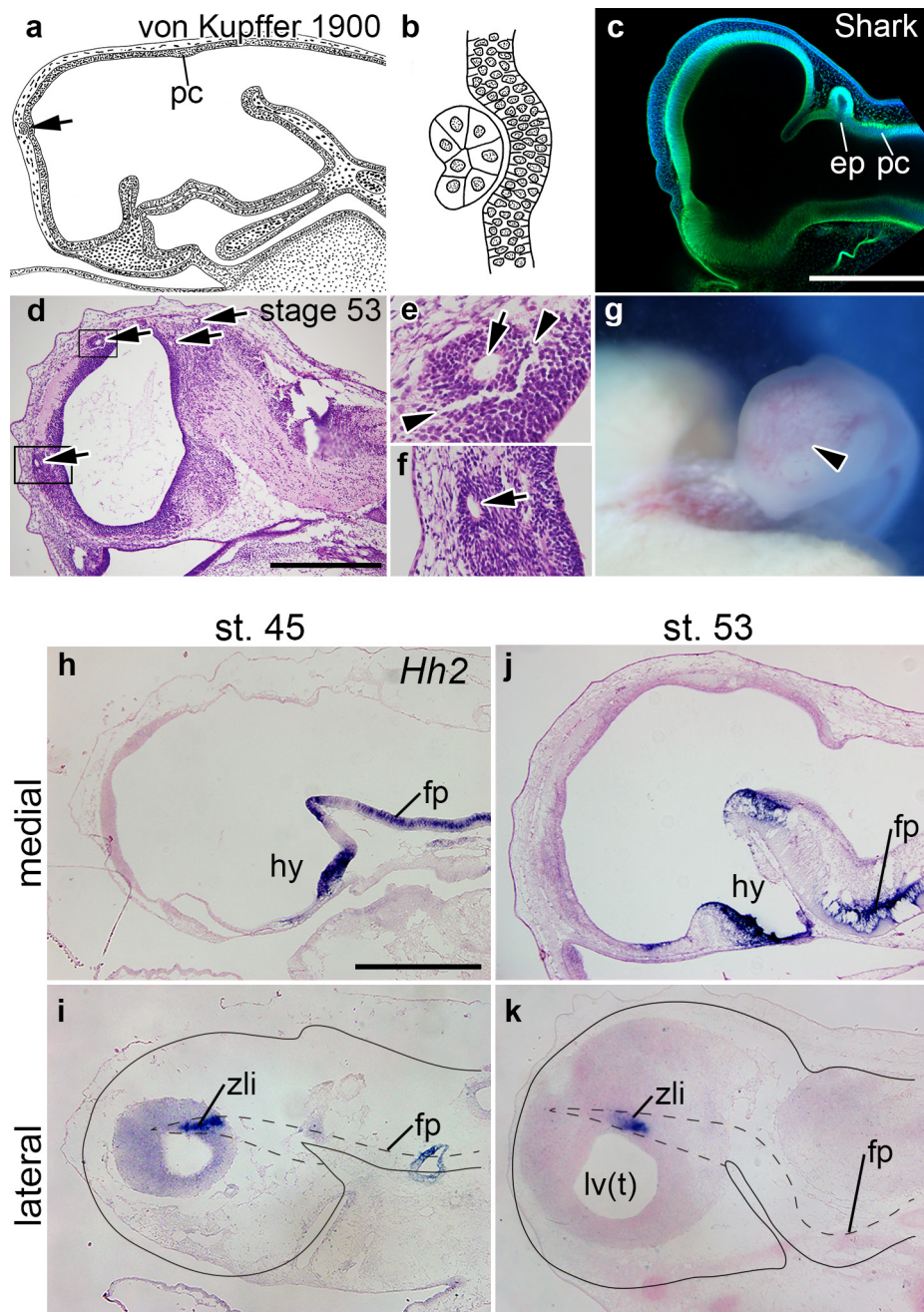
**Extended Data Figure 2 | Adult and embryonic brains of *E. burgeri*, and gene expression and nervous system staining of stage-45 embryo.** a, b, Dorsal view (a) and sagittal section (b) of the adult brain of *E. burgeri*. c–h, *E. burgeri* embryos at Bashford Dean stages 45 (c–e) and 53 (f–h)<sup>17,18</sup>. Dorsal (c, f) and lateral (d, g) views are shown. e, h, Sagittal sections stained with haematoxylin and eosin. Dean stage 53 is very similar to the embryo described in figure 1 in ref. 45. i–t, A stage 45 *E. burgeri*

embryo observed using histological preparations. Dotted lines indicate the presumptive telencephalic border estimated by *EmxB* and *FoxG* expressions<sup>21,22</sup>. At this stage, *Nkx2.1/2.4* and *Hh2* have not been expressed yet in the rostral telencephalon (according to the prosomeric model<sup>3</sup>). cer, cerebrum; met, metencephalon; mes, mesencephalon; ob, olfactory bulb. See Figs 1, 2 and 4 for other abbreviations. Scale bars, 2.0 mm (a–c and f), 1.0 mm (e, h) and 500  $\mu$ m (i, j).



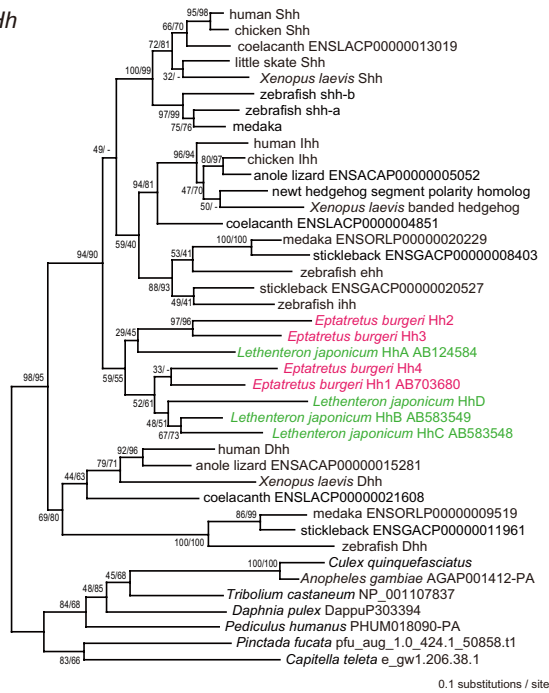
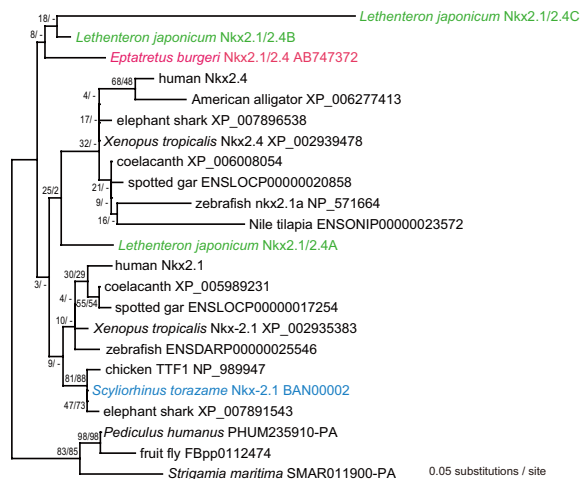
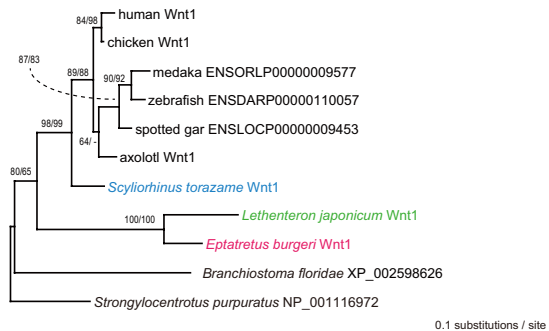
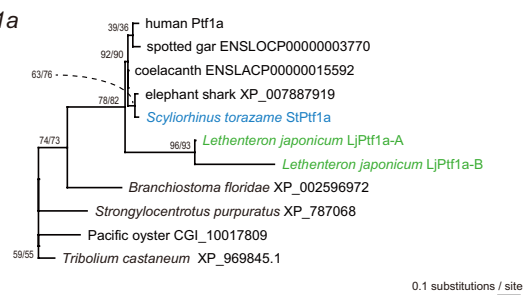
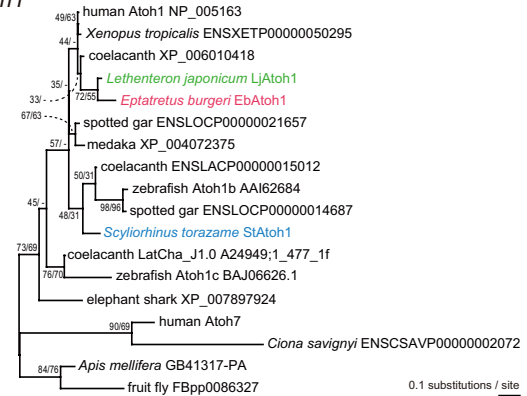
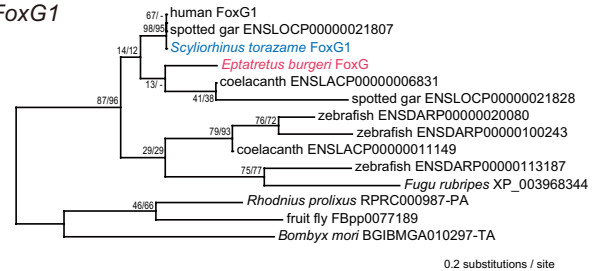
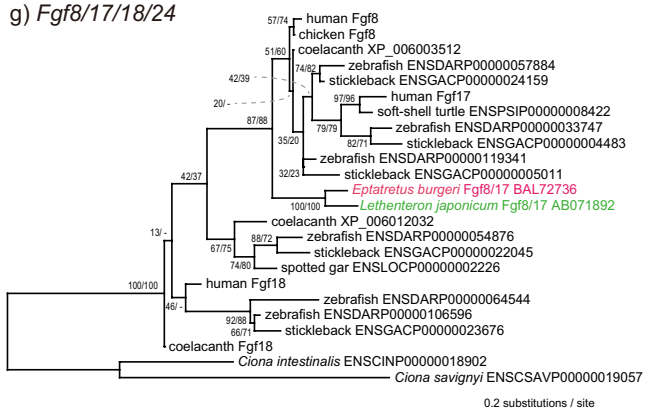


**Extended Data Figure 3 | Identification of the forebrain-midbrain-hindbrain boundaries in early hagfish embryos.** **a-c**, The mid-hindbrain boundary (MHB) of the hagfish brain is assumed to correspond to the posterior expression boundary of *OtxC* (**b**) and focal expression of *Fgf8/17* (see Extended Data Fig. 5) (**c**) that become evident at stages 28–30. **a-d**, On the other hand, the fore-midbrain boundary (FMB) is suggested morphologically (**a**, **b**) and by the ventrocaudal portion of *Nkx2.1/2.4* expression domain (**d**, arrowheads). See Figs 1 and 2 for abbreviations. Scale bar, 200  $\mu$ m.



**Extended Data Figure 4 | Absence of the epiphysis and presence of the ZLI in the developing brain of the hagfish.** **a, b,** The adult hagfish has no epiphysis. A rudimentary epiphysis described previously<sup>14,46</sup> was likely to have been an artefact derived from an inappropriate method of fixation. Although von Kupffer (1900) described an epiphysis-like structure in his illustration (**a, b**, redrawn from ref. 15), he denied its identity as an epiphysis because it is distantly positioned from the posterior commissure (pc). **c,** The position of the epiphysis (ep) in a shark embryo is shown. **d–g,** In our present observation of hagfish embryos also, several

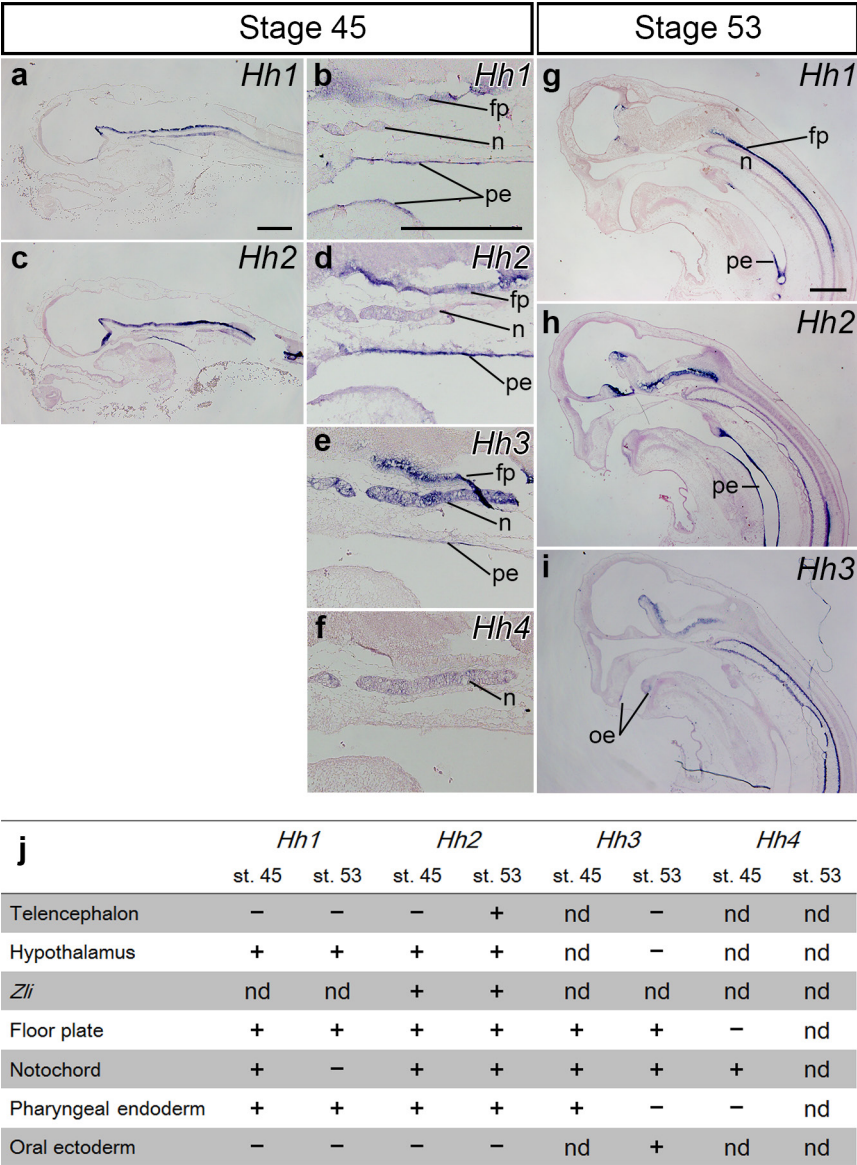
neuroepithelial cysts were found in the forebrain (arrows in **d–f**), which were always associated with the development of blood vessel formation in the brain (arrowhead in **g**). Arrowheads in **e** indicate the crack between the neuroepithelial cyst and the brain tissue caused during the dehydration process, reminiscent of von Kupffer's epiphysis-like structure in **b**. **h–k,** *Hh2* expression in stage-45 (**h, i**) and stage-53 (**j, k**) hagfish embryos. Medial sections (**h, j**) and lateral sections (**i, k**). Dotted lines in **i** and **k** indicate the presumptive ZLI region. See Figs 1–3 for abbreviations. Scale bars, 500 μm.

a) *Hh*b) *Nkx2.1/2.4*c) *Wnt1*d) *Ptf1a*e) *Atoh1*f) *FoxG1*g) *Fgf8/17/18/24*

**Extended Data Figure 5 | Molecular phylogenetic trees of genes identified in this study.** The trees were inferred with the maximum-likelihood method using PhyML3.0 with the JTT+G4 model. Members of individual gene families were collected from public databases, and groupings of the sequences identified in this study into the subfamilies shown in this figure were confirmed by preliminary phylogenetic inferences including other subfamilies. **a**, *Hedgehog* homologues (shape parameter for the gamma distribution  $\alpha = 0.80$ ). A total of 249 amino acid sites unambiguously aligned without any gap were used in the inference. **b**, *Nkx2.1/2.4* homologues (86 sites;  $\alpha = 0.60$ ).

**c**, *Wnt1* homologues (196 sites;  $\alpha = 0.60$ ). **d**, *Ptf1a* homologues (67 sites;  $\alpha = 0.92$ ). **e**, *Atoh1* homologues (60 sites;  $\alpha = 0.74$ ). **f**, *Fgf8/17/18/24* homologues including the hagfish homologue identified previously<sup>19</sup> (74 amino acid sites;  $\alpha = 0.87$ ). **g**, *FoxG1* homologues (115 amino acid sites;  $\alpha = 0.16$ ). Hagfish, lamprey and catshark sequences are shown in red, green and blue, respectively. Support values at nodes are bootstrap probabilities in the maximum-likelihood method and those in the neighbour-joining method (under the above-mentioned substitution model), in order.

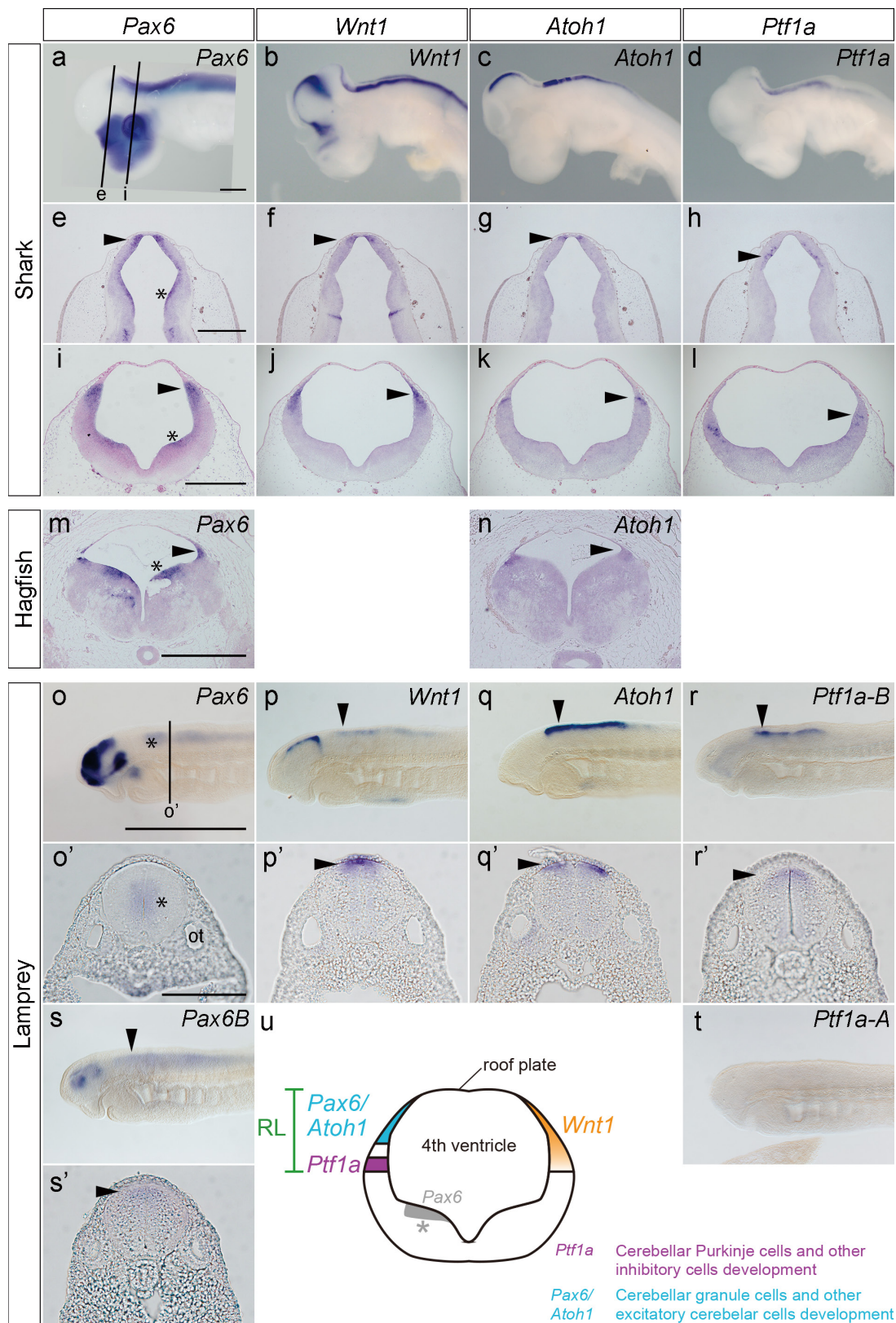




+, present; –, absent; nd, no data

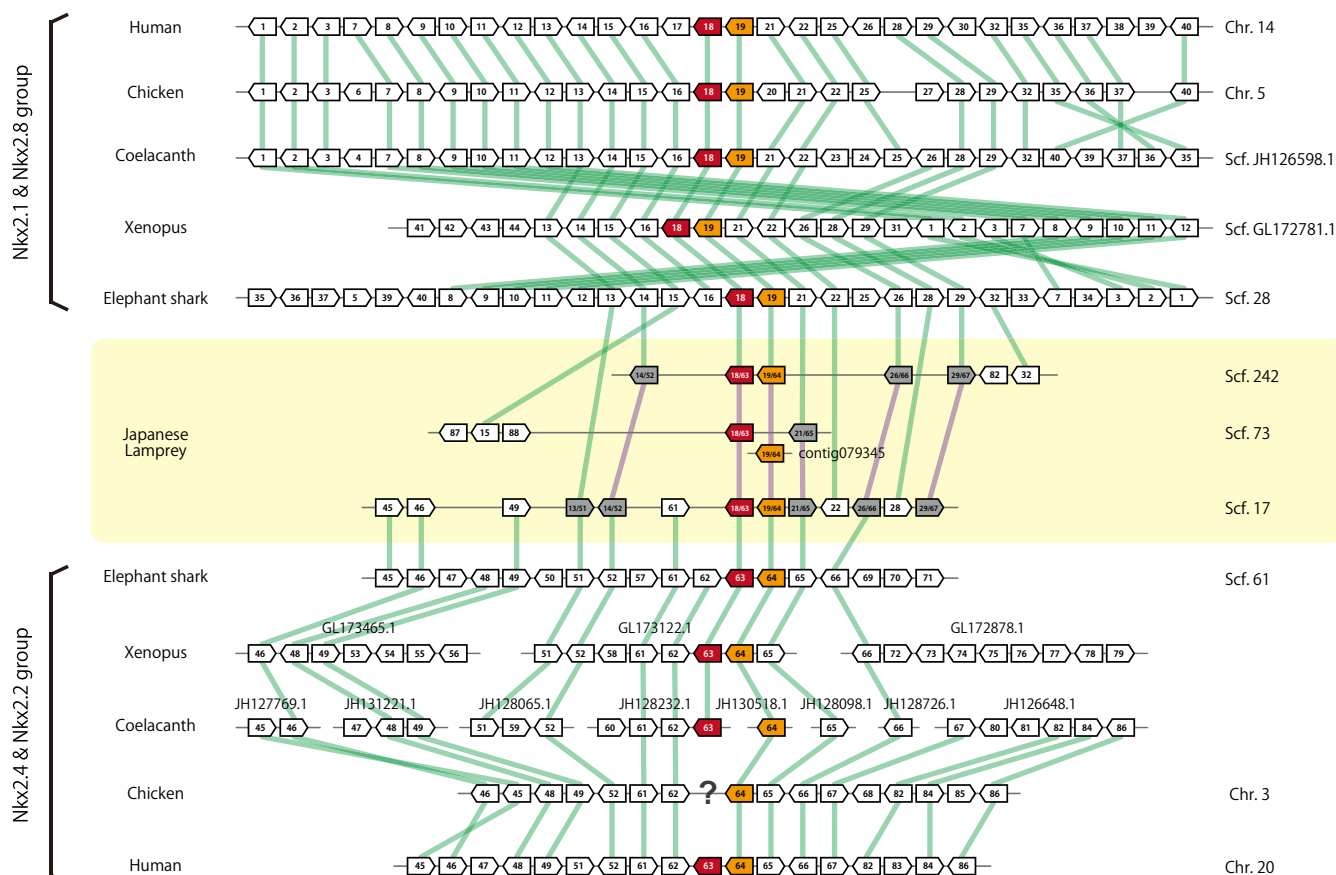
**Extended Data Figure 6 | Expression patterns of hagfish *Hedgehog* genes.** **a–i**, *In situ* hybridization staining of four *Hedgehog* genes (*Hh1–4*) of *E. burgeri* at stage-45 (**a–f**) and stage-53 (**g–i**) embryos. The *Hedgehog* expressions were found in notochord (n), floor plate (fp), pharyngeal endoderm (pe), oral ectoderm (oe) and subregions of the forebrain

(hypothalamus, ZLI and MGE; see Fig. 3). The overall expressions of hagfish *Hh1–4* exhibited patterns similar to those of *Hedgehog* genes in jawed vertebrates, although each *Hh* expression showed slight differences in some domains. **j**, Summary of the expression patterns of *Hedgehog* genes (*Hh1–4*) of *E. burgeri* in stage 45 and stage 53 embryos. Scale bars, 500  $\mu$ m.



**Extended Data Figure 7 | Rhombic lip gene expression patterns in shark, hagfish and lamprey embryos.** Gene expression patterns in stage-27 (whole-mount) and stage-31 (transverse sections) embryos of a catshark (*S. torazame*) (a–l), stage-53 embryo of a hagfish (*E. burgeri*) (m, n), and stage-26 embryo of lamprey (*L. japonicum*) (o–t), stained using probes for the *Pax6*, *Wnt1*, *Atoh1* and *Ptf1a* gene homologues. **u**, Schematic transverse section of the vertebrate rhombic lip showing crucial gene expression patterns based on ref. 27 and the present study.

Lines in **a** indicate the levels of the transverse sections shown in **e–h** (rhombomere 1: the cerebellar primordia) and **i–l** (posterior rhombomeres). The line in **o** indicates the level of the transverse sections shown in (**o'–s'**) (around rhombomere 4). Arrowheads indicate rhombic lip gene expressions. Asterisks indicate non-rhombic lip expressions of *Pax6* through the neural tube. See Figs 1 and 2 for abbreviations. Scale bars, 500  $\mu$ m (**a**, **e**, **i**, **m**, **o**) and 100  $\mu$ m (**o'**).

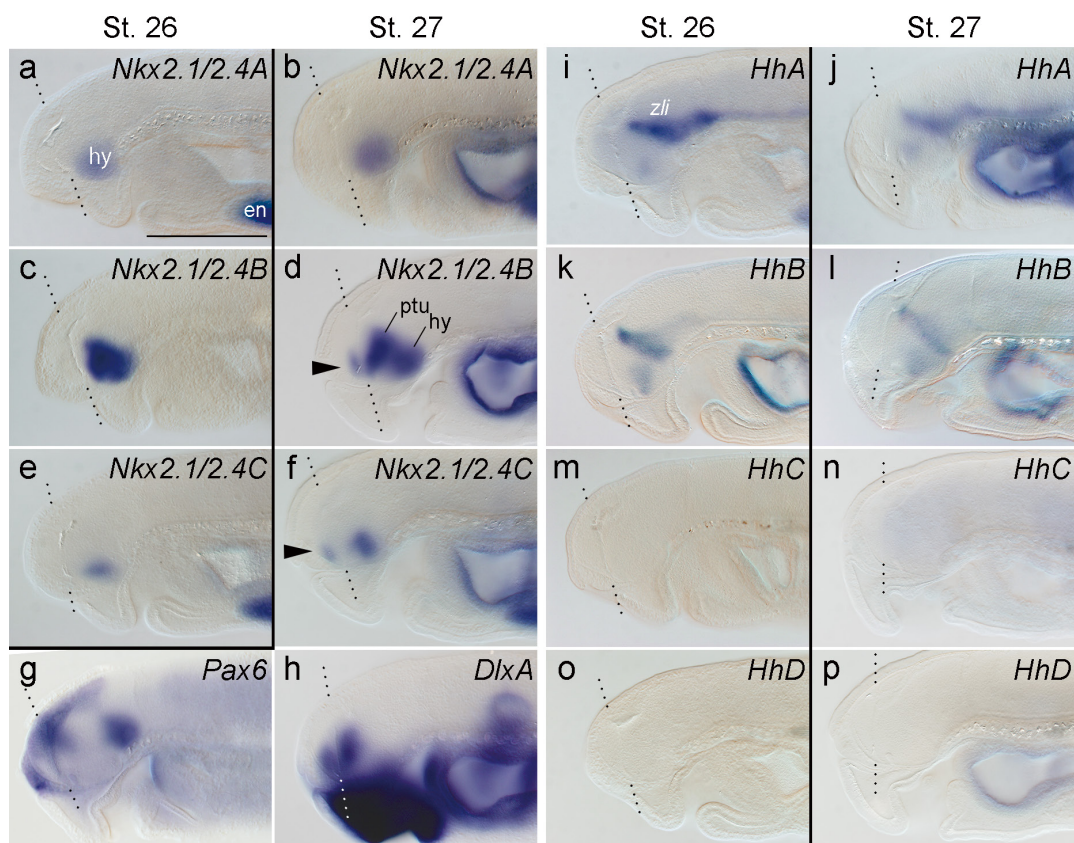


**Extended Data Figure 8 | Synteny conservation between the genomic regions containing *Nkx2.1/2.4* among gnathostomes and *L. japonicum*.**

Numbered boxes represent single protein-coding genes; all the genes used for synteny analysis and the given numbers are listed in Extended Data Table 1. Colours of boxes are as follows: red, *Nkx2-1/2-4* orthologue; orange, *Nkx2-2/2-8* orthologue; grey, a gene showing high sequence

similarity to two gnathostome orthologues as a result of the TBLASTN search. The green line represents orthology between the neighbouring vertebrate genomes, whereas the purple line indicates a paralogous relationship of each gene among the lamprey genome scaffolds. At present, there are no descriptions of chicken *NKX2-4* (nor in birds) and we have not been able to find it. This putative lack is marked by a question mark.





q	<i>Nkx2.1/2.4A</i>	<i>Nkx2.1/2.4B</i>	<i>Nkx2.1/2.4C</i>
Telencephalon	–	+	+
Hypothalamus	+	+	+
Posterior tuberculum	–	+	–
Endostyle	+	+	+
		(weak)	

	<i>HhA</i>	<i>HhB</i>	<i>HhC</i>	<i>HhD</i>
Telencephalon	–	–	–	–
Hypothalamus	+	+	–	–
<i>Zli</i>	+	+	–	–
Floor plate	+	+	–	–
Notochord	+	+	–	–
Pharyngeal endoderm	+	+	–	–
Oral ectoderm	–	–	–	–
Endostyle	+	–	–	–

**Extended Data Figure 9 | Telencephalic gene expression patterns in lamprey embryos.** a–p, Gene expression patterns in *L. japonicum* stage-26 and -27 (ref. 33) embryos, stained using probes for *Nkx2.1/2.4A* (a, b), *Nkx2.1/2.4B* (c, d), *Nkx2.1/2.4C* (e, f), *Pax6* (g), *DlxA* (h), *HhA* (i, j), *HhB* (k, l), *HhC* (m, n) and *HhD* (o, p). Dotted lines indicate the telencephalic border distinguished by the anterior intraencephalic sulcus<sup>25</sup>. The expression domains of *Pax6* (g) and *DlxA* (h) in the telencephalon pallial and subpallial subdivisions, respectively. Although the previously

reported *Nkx2.1/2.4A* is not expressed in the telencephalon<sup>36</sup>, *Nkx2.1/2.4B* and *Nkx2.1/2.4C* are expressed in the rostral telencephalon (arrowheads), suggesting the presence of the MGE in the lamprey. We did not detect expression of any *Hedgehog* genes in the rostral telencephalon<sup>25</sup> (i–p). q, Summary of the expression patterns of *Nkx2.1/2.4* and *Hedgehog* genes of *L. japonicum* at embryonic stages 17–30 examined by whole-mount *in situ* hybridization. en, endostyle; ptu, posterior tuberculum<sup>47</sup>. See Figs 1 and 2 for other abbreviations. Scale bar, 200 μm.

Extended Data Table 1 | List of genes used in the synteny conservation analysis

Nkx2.1 & Nkx2.8 group		Nkx2.4 & Nkx2.2 group	
Number	Gene name	Number	Gene name
1	SNX6	45	SLC24A3
2	CFL2	46	RIN2
3	BAZ1A	47	NAA20
4	ENSLACG00000017921	48	CRNKL1
5	SINCAMG00000016188	49	CFAP61
6	ENSGALG00000010034	50	GTF2IRD2
7	SRP54	51	INSM1
8	FAM177A1	52	RALGAPA2
9	PPP2R3C	53	ATP2A1
10	KIAA0391	54	ENSXETG00000032786
11	PSMA6	55	TCTN3
12	NFKBIA	56	PPP1CA
13	INSM2	57	SINCAMG00000008965
14	RALGAPA1	58	ENSXETG00000009758
15	BRMS1L	59	ENSLACG00000009798
16	MBIP	60	ENSLACG00000001348
17	SFTA3	61	KIZ
18	<b>NKX2.1</b>	62	XRN2
19	<b>NKX2.8</b>	63	<b>NKX2.4</b>
20	ENSGALG00000010113	64	<b>NKX2.2</b>
21	PAX9	65	PAX1
22	SLC25A21	66	FOXA2
23	ENSLACG00000014546	67	SSTR4
24	ENSLACG00000014339	68	THBD
25	MIPOL1	69	SINCAMG00000009008
26	FOXA1	70	SFT2D1
27	ENSGALG00000010121	71	PRR18
28	TTC6	72	CFAP36
29	SSTR1	73	SMEK2
30	CLEC14A	74	PNPT1
31	CAPN3	75	EFEMP1
32	SEC23A	76	CCDC85A
33	HRH2	77	VRK2
34	SINCAMG00000016299	78	FANCL
35	GEMIN2	79	ENSXETG00000020482
36	TRAPPC6B	80	ENSLACG00000017923
37	PNN	81	ENSLACG00000017898
38	MIA2	82	CD93
39	CTAGE5	83	NXT1
40	FBXO33	84	GZF1
41	CHST9	85	ENSGALG00000008350
42	ITPKA	86	NAPB
43	ZNF106	87	NUDT18
44	EAPP	88	PRDX5

Each gene number corresponds to the numbers in Extended Data Fig. 8.

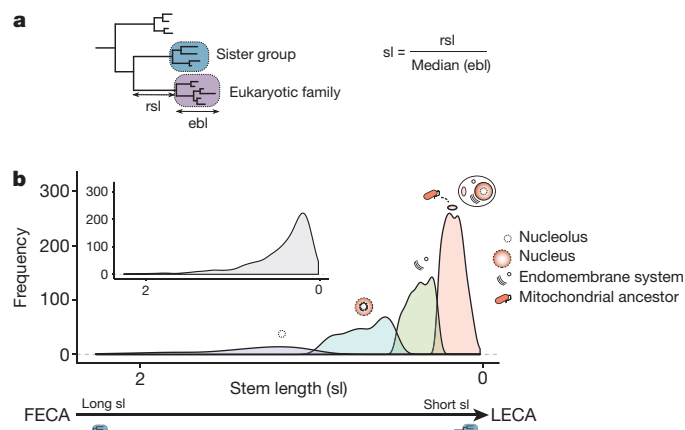
# Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry

Alexandros A. Pittis<sup>1,2</sup> & Toni Gabaldón<sup>1,2,3</sup>

The origin of eukaryotes stands as a major conundrum in biology<sup>1</sup>. Current evidence indicates that the last eukaryotic common ancestor already possessed many eukaryotic hallmarks, including a complex subcellular organization<sup>1–3</sup>. In addition, the lack of evolutionary intermediates challenges the elucidation of the relative order of emergence of eukaryotic traits. Mitochondria are ubiquitous organelles derived from an alphaproteobacterial endosymbiont<sup>4</sup>. Different hypotheses disagree on whether mitochondria were acquired early or late during eukaryogenesis<sup>5</sup>. Similarly, the nature and complexity of the receiving host are debated, with models ranging from a simple prokaryotic host to an already complex proto-eukaryote<sup>1,3,6,7</sup>. Most competing scenarios can be roughly grouped into either mito-early, which consider the driving force of eukaryogenesis to be mitochondrial endosymbiosis into a simple host, or mito-late, which postulate that a significant complexity predated mitochondrial endosymbiosis<sup>3</sup>. Here we provide evidence for late mitochondrial endosymbiosis. We use phylogenomics to directly test whether proto-mitochondrial proteins were acquired earlier or later than other proteins of the last eukaryotic common ancestor. We find that last eukaryotic common ancestor protein families of alphaproteobacterial ancestry and of mitochondrial localization show the shortest phylogenetic distances to their closest prokaryotic relatives, compared with proteins of different prokaryotic origin or cellular localization. Altogether, our results shed new light on a long-standing question and provide compelling support for the late acquisition of mitochondria into a host that already had a proteome of chimaeric phylogenetic origin. We argue that mitochondrial endosymbiosis was one of the ultimate steps in eukaryogenesis and that it provided the definitive selective advantage to mitochondria-bearing eukaryotes over less complex forms.

Previous analyses infer a last eukaryotic common ancestor (LECA) proteome of diverse phylogenetic origin<sup>1,8</sup>. Notably, only a fraction of the proteins of bacterial descent can be traced back to Alphaproteobacteria, the group from which mitochondria originated<sup>4</sup>. Attempts to explain alternative bacterial signals in LECA range from invoking horizontal gene transfer (HGT), phylogenetic noise or additional symbiotic partners<sup>9,10</sup>, including the possibility that part of this diversity could have already been present in the putative archaeal host<sup>11</sup>. Resolving whether LECA proteins of bacterial descent were acquired in bulk is key to testing competing eukaryogenesis models. Here, we set out to assess whether the LECA proteins with alphaproteobacterial ancestry show distinct patterns in terms of their current cellular localizations, and evolutionary distances to their closest ancestors, compared with LECA proteins of other descent. For this, we surveyed the phylogenetic signal of inferred LECA proteomes (see Methods). First, the likely phylogenetic origin of each LECA family was assessed by evaluating the taxonomic distribution of prokaryotic sequences present in its closest neighbouring tree partition (see Methods and Fig. 1a). We then established a measure of phylogenetic distance for the branch subtending the

LECA family and connecting it to the last prokaryotic ancestor shared with its closest prokaryotic relatives (raw stem length; Fig. 1a). Branch lengths indicate the number of inferred substitutions per site, which reflect both divergence time and evolutionary rate. To disentangle time from rates, which may vary across families, we normalized the raw stem length by taking into account the median of the branch lengths within the LECA family (see Methods for further details). We used this measurement (hereafter referred to as stem length) as a proxy for the phylogenetic distance between a given LECA protein family and its last shared ancestor with prokaryotes. Competing mito-early and mito-late hypotheses naturally differ in their expectations of stem lengths for proteins of proto-mitochondrial origin compared with those of other putative origins. In a simple fusion model, with the proto-mitochondrion contributing most of the bacterial component, one would expect



**Figure 1 | Stem length analysis.** **a**, Schematic representation of the inference of the phylogenetic origin of LECA groups and the measured phylogenetic distances. First monophyletic groups of eukaryotic proteins that passed the required thresholds were considered as protein families present in LECA (purple box). The taxonomic range of the proteins present in the closest neighbouring tree partition (sister group, blue box) was used to define the putative phylogenetic origin of the LECA family. Distance to the common ancestor with the closest prokaryotic neighbouring group was measured (raw stem length, rsl) and normalized (stem length, sl) by dividing it by the median of the distances from the eukaryotic terminal nodes to the last common ancestor of all eukaryotic sequences (eukaryotic branch length, ebl). **b**, Subpopulation distributions within the overall stem length distribution (inset) as defined by a mixture model and the expectation-maximization algorithm. The four subpopulations/components are over-represented in different prokaryotic phylogenetic groups of origin, Gene Ontology (GO) and clusters of orthologous groups (COGs) functional category annotations (see text, Table 1 and Supplementary Tables 1 and 2). On top of these components, we represent the cellular localizations for which each family class is enriched. FECA, first eukaryotic common ancestor.

<sup>1</sup>Bioinformatics and Genomics Programme, Centre for Genomic Regulation (CRG), Carrer del Dr Aiguader, 88, 08003 Barcelona, Spain. <sup>2</sup>Departament de Ciències Experimentals I de La Salut, Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. <sup>3</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluís Companys 23, 08010 Barcelona, Spain.



**Table 1 | Over-represented phylogenetic origins, GO terms and functional categories in the different components**

Component	Size	Phylogenetic origin			Cellular localization			Cellular function		
		Group	N	P	GO cellular component	N	P	Functional category	N	P
1	452	Bacteria	388	$<1 \times 10^{-6}$	Mitochondrion	150	$<1 \times 10^{-6}$	Amino-acid transport and metabolism	72	$1.8 \times 10^{-4}$
		Alphaproteobacteria	49	$1.1 \times 10^{-4}$				Energy production and conversion	45	$1.6 \times 10^{-2}$
		Chlamydiae/Verrucomicrobia group	19	$1.4 \times 10^{-2}$				Coenzyme transport and metabolism	29	$4.9 \times 10^{-2}$
		Deltaproteobacteria	29	$2.0 \times 10^{-2}$						
2	284	-			Endoplasmic reticulum	32	$3.5 \times 10^{-3}$	Carbohydrate transport and metabolism	28	$3.7 \times 10^{-2}$
					Golgi apparatus	11	$4.1 \times 10^{-2}$			
					Extracellular space	8	$1.4 \times 10^{-2}$			
3	234	Archaea	80	$<1 \times 10^{-6}$	Nucleoplasm	13	$2.7 \times 10^{-3}$	Replication, recombination and repair	24	$6.1 \times 10^{-4}$
		Euryarchaeota	30	$1.3 \times 10^{-4}$	Nucleus	80	$5.9 \times 10^{-3}$	Translation, ribosomal structure and biogenesis	46	$6.6 \times 10^{-3}$
		Crenarchaeota	15	$3.4 \times 10^{-3}$	Chromosome	14	$7.4 \times 10^{-3}$	Transcription	10	$4.9 \times 10^{-2}$
		Korarchaeota	7	$1.2 \times 10^{-2}$	Nuclear chromosome	9	$2.4 \times 10^{-2}$			
		Actinobacteria	16	$2.7 \times 10^{-2}$	Nucleolus	19	$2.5 \times 10^{-2}$			
					Protein complex	46	$2.9 \times 10^{-2}$			
4	94	Archaea	41	$<1 \times 10^{-6}$	Ribosome	24	$<1 \times 10^{-6}$	Translation, ribosomal structure and biogenesis	36	$<1 \times 10^{-6}$
		Thaumarchaeota	8	$4.9 \times 10^{-4}$	Cytosol	39	$<1 \times 10^{-6}$			
		Euryarchaeota	16	$1.4 \times 10^{-3}$	Organelle	70	$1.7 \times 10^{-2}$			
		Crenarchaeota	7	$2.7 \times 10^{-2}$	Nucleolus	10	$1.4 \times 10^{-2}$			

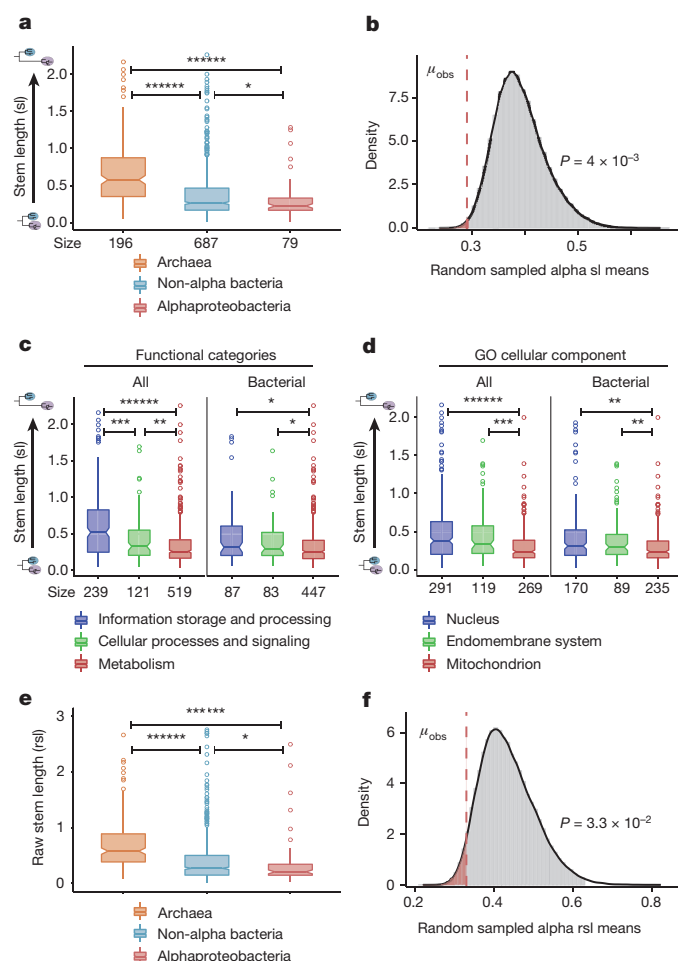
N, number of LECA families per term, in each component. P values  $<10^{-6}$  reflect value 0 in  $10^6$  permutations.

stem lengths of bacterial-derived proteins to be similar. In contrast, significant differences would be predicted by models involving different waves of gene acquisition. We assessed differences in stem length, protein function and subcellular localization across 1,078 LECA families of different origins.

We first used an unsupervised approach to assess whether the distribution of stem lengths in LECA families was homogeneous. By using the expectation-maximization algorithm<sup>12</sup> to fit observed data to a mixture model, we inferred four distinct underlying distributions (Fig. 1b), each containing a subset of LECA families. We asked whether each underlying distribution contained an enrichment of protein families with (1) a particular taxonomic origin, (2) a particular subcellular localization or (3) a particular functional category. Notably, we found that the first component (shortest stems) was enriched in families with bacterial origins (most particularly alphaproteobacterial), mitochondrial localization and energy production (see Table 1). In contrast, the two components with the longest stems (third and fourth) were enriched in families of archaeal and actinobacterial origins, and in annotations related to the nucleus and ribosomes (Fig. 1b and Table 1). The second component showed no enrichment in any ancestry, but a significant enrichment in endomembrane system localization. The above results are only consistent with mito-late models, with the archaeal contributions to eukaryotes, mainly associated with nuclear structures and genes related to informational processes (replication, transcription, translation), being more ancient; with the prokaryotic proteome of the endomembrane system being integrated later; and with the alphaproteobacterial contribution, associated with mitochondria and energy production, appearing later than other bacterial components.

We tested this hypothesis more directly by grouping the LECA families by their inferred phylogenetic origin, and by their functional and subcellular localization annotations, and then testing whether their respective stem lengths were significantly different (Fig. 2a and Extended Data Fig. 1a–c). Overall, LECA families of bacterial origins have significantly shorter stems than families of archaeal origin ( $P = 1.38 \times 10^{-25}$ , two-sided Mann–Whitney U-test). Importantly, eukaryotic families of alphaproteobacterial descent

showed the shortest stems, together with families pointing to the Verrucomicrobia/Chlamydiales group. These lengths were significantly smaller than those found in LECA families of different bacterial origins ( $P = 4.4 \times 10^{-2}$ ). When grouping LECA families according to their functional annotations, we found that those involved in informational processes had the longest stems, followed by those involved in cellular and signalling processes, with families involved in metabolic processes showing the shortest stems (Fig. 2c and Extended Data Fig. 1c). Next, we asked whether LECA families predominantly present in distinct subcellular compartments showed differences in terms of phylogenetic origins and stem lengths. Consistent with the above results, nuclear protein families had the longest stems, followed by those involved in the endomembrane system, and finally mitochondrial proteins tended to have the shortest stems (Fig. 2d and Extended Data Fig. 1d). The fact that both function and evolutionary origin correlate with stem length raises the need to disentangle the contribution of each of these factors. Our normalization assumes proportional (not necessarily constant) evolutionary rates in branches preceding and post-dating LECA, which both correspond to periods where the given protein had been incorporated into the host. Large shifts in evolutionary rates between the stem and post-LECA phases may have differentially impacted families depending on their function, leading to the observed differences mentioned above. However, our results are independent of the normalization, as shown in comparisons using the raw stem lengths (Fig. 2e, f). Furthermore, in matched comparisons, families of similar function, selection pressure, number of protein–protein interactions or expression levels but different origins show differences in stem lengths (Supplementary Information section 1 and Extended Data Fig. 2). Thus, phylogenetic origin, and not function, is the main driver of observed differences in stem lengths. To independently validate our approach, we assessed the relative timing of the acquisition of plastids, a type of organelle whose origin from cyanobacteria subsequent to mitochondrial endosymbiosis is uncontroversial. Consistently, cyanobacterial-derived families had significantly shorter stem lengths than alphaproteobacterial-derived families, thereby further supporting our approach (Supplementary Information section 2 and Extended Data Fig. 3).



**Figure 2 | Phylogenetic distance profiles.** **a–d**, Profiles of different prokaryotic sources (**a**, **b**), cellular functions (**c**) and cellular components (**d**). The lower and upper box limits in **a**, **c** and **d** correspond to the first and third quartiles (25th and 75th percentiles). **a**, Box plot comparing stem length distributions in LECA families with archaeal, non-alpha bacterial and alphaproteobacterial sister-groups. Numbers on the x axis indicate the number of families included in each class. Symbols indicate the *P* values obtained from a two-sided Mann–Whitney *U*-test for the indicated comparisons as follows: \* $P \leq 5 \times 10^{-2}$ ; \*\* $P \leq 1 \times 10^{-2}$ ; \*\*\* $P \leq 1 \times 10^{-3}$ ; \*\*\*\* $P \leq 1 \times 10^{-4}$ ; \*\*\*\*\* $P < 1 \times 10^{-6}$ . **b**, The observed mean ( $\mu_{\text{obs}}$ ) stem length of alphaproteobacterial values compared with the random sampling distribution of means, under the null hypothesis that families of different bacterial origins do not show differences in stem lengths. The *P* value is the probability that the mean would be at least as extreme as the observed, if the null hypothesis were true. The dashed line and the shaded area under the density plot correspond to the one-sided *P* value of the test (indicated next to the figure). **c**, **d**, Box plots of stem length distributions in LECA families of different COG functional categories (**c**) and GO localizations (**d**), when considering all LECA families (All), or only those of bacterial descent (Bacterial). Other symbols as in **a**. **e**, **f**, The results obtained in **a** and **b** are consistent when using raw stem lengths, indicating that the relative differences in stem lengths are not driven by differences in the rates of evolution within extant eukaryotes (ebl).

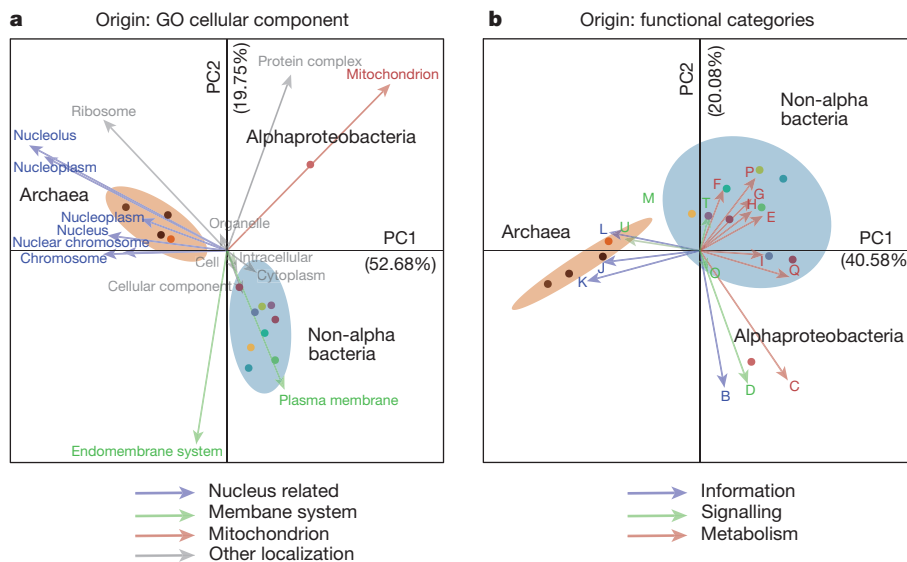
We next tested the robustness of our results with different LECA data sets, sequence sampling and phylogenetic methods (see Supplementary Information sections 3–5 and Extended Data Fig. 4–6). Additional controls (Supplementary Information sections 4 and 5 and Extended Data Fig. 6) showed that HGT alone cannot explain the observed signal from non-alphaproteobacterial bacteria, and discarded the possibility that shorter stem lengths in alphaproteobacterial-derived families resulted only from specific functional classes, or from those affiliated to Rickettsiales, whose

specific clustering to mitochondrial proteins has been considered artefactual<sup>13</sup>. Finally, we included data from the recently identified lokiarchaeon clade in our analysis<sup>11</sup>. Even though we found that LECA families of inferred lokiarchaeal origin had stems larger than those of bacterial-derived families, they did show the shortest stems among archaeal-derived proteins, thereby providing additional support that there is a close association of this clade to eukaryotes (Supplementary Information section 6 and Extended Data Fig. 7).

To gain further insight into the functionality and localization of the LECA families of different phylogenetic origins, we used correspondence analysis to visualize associations among these variables, and permutation tests to assess the statistical significance (see Methods, Fig. 3 and Extended Data Fig. 8). We found that alphaproteobacterial-derived genes tend to associate with mitochondria ( $P \leq 10^{-6}$ , permutation test with  $10^6$  randomizations), whereas archaeal-derived families do so with the nucleus. Perhaps more unexpectedly, we found that LECA families of bacterial descent, except for Alphaproteobacteria, showed a clearly distinct pattern, being predominantly associated with endomembrane related compartments (Fig. 3b and Extended Data Fig. 8b, d). Consistent results were obtained when correlations between evolutionary origins and functional categories were evaluated (Fig. 3a and Extended Data Fig. 8a, c). In particular, the alphaproteobacterial component showed a unique correlation with energy production ( $P < 10^{-6}$ ). This result is not consistent with scenarios in which most of the bacterial components in LECA are assumed to originate from the alphaproteobacterial endosymbiont, because in this case a higher functional coherence would be expected among them. These results also reinforce the idea that, despite substantial subcellular re-targeting and functional diversification, the proto-mitochondrial-derived fraction of the eukaryotic proteome retains a tendency to be mitochondrial localized<sup>14</sup>. Interestingly, alphaproteobacterial-derived families of mitochondrial localization have shorter stem lengths than mitochondrial families of different origins ( $P = 6 \times 10^{-3}$ ), which indicates re-targeting to the newly formed organelle.

Altogether, our results provide compelling support for a late acquisition of mitochondria, as proposed by several eukaryogenesis models<sup>5</sup>. Specifically, our data suggest that most of the bacterial component of LECA, with origins other than alphaproteobacteria, was acquired earlier and mostly contributed to compartments other than the mitochondrion or the nucleus, and to processes besides energy production. We have shown that this pattern cannot be entirely explained by massive HGT to the proto-mitochondrial ancestor. This implies that these proteins were acquired by the host genome before mitochondrial acquisition. Thus, the host that engulfed the mitochondrion was already a complex cell, whose genome already harboured pathways and processes of diverse bacterial origins. Given the heterogeneity of these alternative bacterial origins, no simple model can explain this component. Serial symbiotic associations with different partners, the existence of prokaryotic consortia or gradual waves of HGT to the host genome before mitochondrial endosymbiosis could all explain such chimaerism. Finally, the archaeal-derived component has the longest stems and the strongest association with the nucleus, consistent with the idea that eukaryotes have rooted from within archaea, and that the nucleus is of archaeal origin. Our results are compatible with either a complex proto-eukaryotic host or a complex archaeal host already harbouring many pathways of bacterial origin<sup>15</sup>. In either case, mitochondrial engulfment marked an end to massive bacterial HGT in LECA and the start of the diversification of extant eukaryotic lineages. We argue that mitochondrial endosymbiosis was indeed a crucial late step in eukaryogenesis, which brought about the definitive selective advantage that facilitated the dominance and radiation of the eukaryotic groups that have survived to the present day.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.



**Figure 3 | Correspondence of different LECA components with different cellular localizations and functions.** **a, b**, Correspondence analysis symmetrical biplots showing differences between the localizations (**a**) and functions (**b**) of the families of various phylogenetic origins. In both cases, the first principal components, accounting for the largest percentage of variance explained, clearly separate the bacterial and archaeal (brown ellipse) eukaryotic origins, while the second components separate the alphaproteobacterial (red dot) from the other bacterial origins (cyan ellipse). The numbers next to the principal axes (PC1, PC2) show

the percentage of the total variance explained by each component. Both columns (functions or localizations) and rows (phylogenetic origins) are in principal coordinates. The colours of the arrows, cellular localizations (left) and functional categories (right) correspond to the categories and localizations of Fig. 2c, d accordingly (see Methods). If a term cannot be categorized as above, the colour is grey. Dots are coloured according to the phylogenetic origin of the group as in Extended Data Fig. 1a (see also extended version of this in Extended Data Fig. 8).

Received 18 September; accepted 21 December 2015.

Published online 3 February 2016.

- Koonin, E. V. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* **11**, 209 (2010).
- Embley, T. M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630 (2006).
- Koumandou, V. L. *et al.* Molecular paleontology and complexity in the last eukaryotic common ancestor. *Crit. Rev. Biochem. Mol. Biol.* **48**, 373–396 (2013).
- Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).
- Poole, A. M. & Gribaldo, S. Eukaryotic origins: how and when was the mitochondrion acquired? *Cold Spring Harb. Perspect. Biol.* **6**, a015990 (2014).
- Martijn, J. & Ettema, T. J. G. From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochem. Soc. Trans.* **41**, 451–457 (2013).
- Lester, L., Meade, A. & Pagel, M. The slow road to the eukaryotic genome. *BioEssays* **28**, 57–64 (2006).
- Rochette, N. C., Brochier-Armanet, C. & Gouy, M. Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. *Mol. Biol. Evol.* **31**, 832–845 (2014).
- Thiergart, T., Landan, G., Schenk, M., Dagan, T. & Martin, W. F. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biol. Evol.* **4**, 466–485 (2012).
- Ku, C. *et al.* Endosymbiotic gene transfer from prokaryotic pangenomes: inherited chimerism in eukaryotes. *Proc. Natl Acad. Sci. USA* **112**, 10139–10146 (2015).
- Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).

- Do, C. B. & Batzoglou, S. What is the expectation maximization algorithm? *Nature Biotechnol.* **26**, 897–899 (2008).
- Esser, C. *et al.* A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21**, 1643–1660 (2004).
- Gabaldón, T. & Huynen, M. A. Shaping the mitochondrial proteome. *Biochim. Biophys. Acta* **1659**, 212–220 (2004).
- Koonin, E. V. & Yutin, N. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb. Perspect. Biol.* **6**, a016188 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** T.G. group research is funded in part by a grant from the Spanish Ministry of Economy and Competitiveness (BIO2012-37161), a grant from the European Union FP7 FP7-PEOPLE-2013-ITN-606786 and a grant from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC (Grant Agreement number ERC-2012-StG-310325).

**Author Contributions** A.A.P. and T.G. conceived the study. A.A.P. performed the computational analyses. A.A.P. and T.G. analysed and interpreted the data. A.A.P. and T.G. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.G. ([tgabaldon@crg.es](mailto:tgabaldon@crg.es)).



## METHODS

No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

**Sequence data.** The sequences of proteins encoded by 3,686 fully-sequenced genomes of eukaryotes (238), Bacteria (3,318) and Archaea (130), as well as the 192,421 non-supervised orthologous groups (NOGs) and COGs corresponding to the broadest taxonomic level (last universal common ancestor, LUCA), were downloaded from eggNOG version 4.0 (ref. 16); hereafter NOGs/COGs will be referred to as orthologous groups, indistinctively. In total, 11,504 orthologous groups containing 4,323,066 sequences both from eukaryotic and from prokaryotic species were considered. For the analysis including the recently sequenced member of *Lokiarchaeota*<sup>11</sup>, the 5,384 protein coding sequences of the archaeon *Loki* were downloaded as of 7 May 2015 from the Protein database of NCBI (<http://www.ncbi.nlm.nih.gov/protein/>) under the taxonomy identifier 1538547.

**Taxonomy-based sequence sub-sampling.** To reduce data redundancy and obtain a more balanced representation of different eukaryotic families, the initial data set was sub-sampled using taxonomic criteria. We selected 37 eukaryotic species, covering all main eukaryotic subdivisions present in EggNOG version 4 (unikonts, Archaeplastida, Chromalveolates, Excavates), emphasizing model species for which better genomes with experimental annotations were available (Supplementary Table 1). The selected set comprises 18 unikonts (16 Opisthokonta and 2 Amoebozoa), 6 Archaeplastida (5 Viridiplantae and 1 Rhodophyta), 8 Chromalveolates (5 Alveolata, 3 Stramenopiles) and 5 Excavates (2 Euglenozoa, 1 Fornicata, 1 Parabasalia and 1 Heterolobosea). Similarly, for the prokaryotic genomes, we defined 692 levels based on taxonomic criteria. This set represents all 681 prokaryotic genera present in EggNOG version 4 and 11 groups in which the 'genus' rank is not assigned ('no rank'). Genomes with non-informative taxonomic assignments, including the words 'environmental' and 'unclassified', were not considered. For each of the orthologous groups, we randomly sampled one sequence from each of the 729 taxonomic levels defined (37 eukaryotic species plus 692 prokaryotic levels).

**Phylogenetic analysis and identification of LECA families.** The detection of LECA families (that is, groups of related eukaryotic sequences that are inferred to be derived from LECA) was done in two steps. First, maximum likelihood trees were computed using a fast approach. For this we first built alignments of the 8,188 filtered orthologous groups using MAFFT version 6.861b<sup>17</sup> and the -auto parameter. These alignments were trimmed using trimAl version 1.4 (ref. 18) with a gap score cutoff of 0.01. Then, maximum likelihood phylogenetic trees were reconstructed using FastTree 2.1.7 (ref. 19) and the WAG evolutionary model (-wag). These trees were inspected to identify monophyletic groups of three or more eukaryotic sequences, corresponding to eukaryotic protein families. Similarly to previous studies<sup>8</sup>, eukaryotic sequences within one orthologous group were not considered *a priori* monophyletic, as the same group could comprise different eukaryotic groups derived from ancestral duplications subsequent to LUCA but preceding LECA (see also ref. 20). This resulted in the identification of multiple eukaryotic LECA families in some orthologous groups.

In the subsequent step we performed a second phylogenetic analysis of the identified eukaryotic LECA families. For this we considered only the sequences in the given eukaryotic family and all the prokaryotic sequences in the tree, and used a more accurate phylogenetic approach. We used a similar pipeline to that described in ref. 21. In brief, multiple sequence alignments using three different aligners, MUSCLE version 3.8.31 (ref. 22), MAFFT version 6.861b<sup>17</sup> and DIALIGN-TX 1.0.2 (ref. 23), were performed in forward and reverse orientation. The six resulting alignments were combined with M-COFFEE version 8.80 (ref. 24) into a maximal-consensus alignment, which was trimmed using trimAl version 1.4 (ref. 18) with a gap score cutoff of 0.01. For each sequence alignment, the best-fit evolutionary model selection was done before phylogenetic inference using ProtTest version 3 (ref. 25). In each case three different evolutionary models were tested (JTT, WAG, LG). The model best fitting the data was determined by comparing the likelihood of all models according to the Akaike information criterion. Finally, an maximum likelihood tree was inferred with RAxML version 8.0.22 (ref. 26) using the best-fitting model and a discrete gamma-distribution model with four rate categories plus invariant positions. The gamma parameter and the fraction of invariant positions were estimated from the data. SH-like branch support values were computed using RAxML version 8.0.22. Only the eukaryotic families whose monophyly was also recovered in this second phylogenetic step, and for which the support value of the branch between this clade and the prokaryotic sister clade was higher than 0.5, were further considered in the analysis. For the phylogenetic analysis, the execution of the different phylogenetic workflows was done using the bioinformatics tool ETE version 2.3 (ref. 27) as environment in the single gene tree execution mode.

**Detection of eukaryotic families present in LECA.** Our workflow provided us with a flexible framework for evaluating the effect on the final outcome of

different definitions of LECA. Results using alternative criteria are discussed in Supplementary Information section 1. Similarly to previous analyses<sup>8</sup>, a eukaryotic family was inferred as being derived from LECA on the basis of its presence in different major eukaryotic groups. In particular, the requisites for inclusion in LECA are similar to the one used in ref. 8, but with some important differences. For instance, the criteria used in ref. 8 could be met by genes present only in Archaeplastida and Chromalveolates, a pattern that suggests genes are acquired in Chromalveolates through secondary endosymbiosis<sup>28</sup>. Our criteria required the presence of sequences both from unikonts and from at least one of the other groups among bikonts (Archaeplastida, Chromalveolates and Excavates; see also Extended Data Fig. 4a). This procedure rendered 1,078 families, 433 of which were present in all 4 groups and 323 in at least three groups, including unikonts. Upon using more stringent definitions our main results were not affected, but the number of families that could be selected for analysis was significantly reduced (see Extended Data Fig. 4b and Supplementary Information section 3.1).

**Inference of the prokaryotic sister group and phylogenetic origin.** We used a nearest neighbour approach for estimating the prokaryotic affiliation of each LECA group (see Fig. 1a). For that, the phylogenetic trees were first rooted to the prokaryotic sequence that was most distant from the eukaryotic LECA family. Then, the phylogenetic origin of each LECA family was assigned by evaluating the prokaryotic species present in the sister tree partition and using the NCBI taxonomy to define the narrowest taxonomic level that included all prokaryotic species present in that partition. For instance, if only sequences from Alphaproteobacteria and Betaproteobacteria were present in the sister branch, the inferred origin would be 'proteobacteria'. If sequences from any bacterial group(s) were present together with sequences from any archaeal group(s), the group of origin would be considered 'cellular organisms' and so on. Given the hierarchical structure of NCBI taxonomy, this assignment inherited all parent taxonomic levels included within it. For example, a LECA family with an inferred origin in Rickettsiales was also assigned alphaproteobacterial, proteobacterial and bacterial origins.

**Measurement of the phylogenetic distance to the last common prokaryotic ancestor of LECA families: stem lengths.** The branch of interest of each gene family tree is the one connecting the last common ancestor of the LECA family with the common ancestor of this and the nearest prokaryotic sister group to LECA (stem, see Fig. 1a). The length of this branch corresponds to the expected number of substitutions per site in that lineage: that is, the amount of change from the incorporation of the gene into the eukaryotic lineage until LECA. As this also depends on the evolutionary rate of each gene, we normalized the stem length value by dividing it by the median of the branch lengths within the LECA family. We chose the median because of its robustness with respect to extreme outliers (very long branches resulting from fast evolving sequences or phylogenetic artefacts). In the text we refer to this corrected branch length value as stem length. Our rationale for this correction is the following: across families, the time of divergence from LECA is, by definition, the same. Therefore, differences in eukaryotic branch lengths across families are expected to reflect differences in evolutionary rates. By applying this correction, we thus divide by a constant (time from LECA) and a rate, which varies from family to family. This can schematically be expressed by the following relationship:

$$\text{stem length} = \frac{R_s \times T_s}{R_e \times T_e}$$

where  $R_s$ ,  $T_s$  and  $R_e$ ,  $T_e$  are the evolutionary rate ( $R$ ) and divergence time ( $T$ ) of the stem ( $s$ ) and the eukaryotic clade ( $e$ ), respectively. Under the assumption that rates pre- and post-LECA are correlated (that is, not necessarily constant), this normalization compensates for differences in rates in the pre-LECA branches, providing a closer measurement of the divergence time from the prokaryotic ancestor to the LECA. Although we cannot discard that major rate shifts in pre- and post-LECA branches occurred in some cases, we consider it unlikely that they affected in a similar way all proteins of the same phylogenetic origin, regardless of their function; or that they affected in an opposite way proteins with similar function but of different phylogenetic origin. Nevertheless, we performed comparisons using the raw stem lengths as well as with the (normalized) stem lengths.

**LECA family descriptors.** Each LECA family was assigned a phylogenetic origin and a normalized stem length. In addition, they also received the functional (COG functional categories) and GO annotations provided by the eggNOG database. Annotations included functional categories of the corresponding orthologous groups as defined in the COG database, as well as GO cellular component annotations, of which we only considered terms that had experimental evidence codes and that were present in the GO slim generic cut-down version of the GO ontologies. After testing alternative thresholds, GO terms were assigned to the corresponding families if they were present in more than 10% of the sequences in the family considered. For the correspondence analysis (see below), where very rare terms

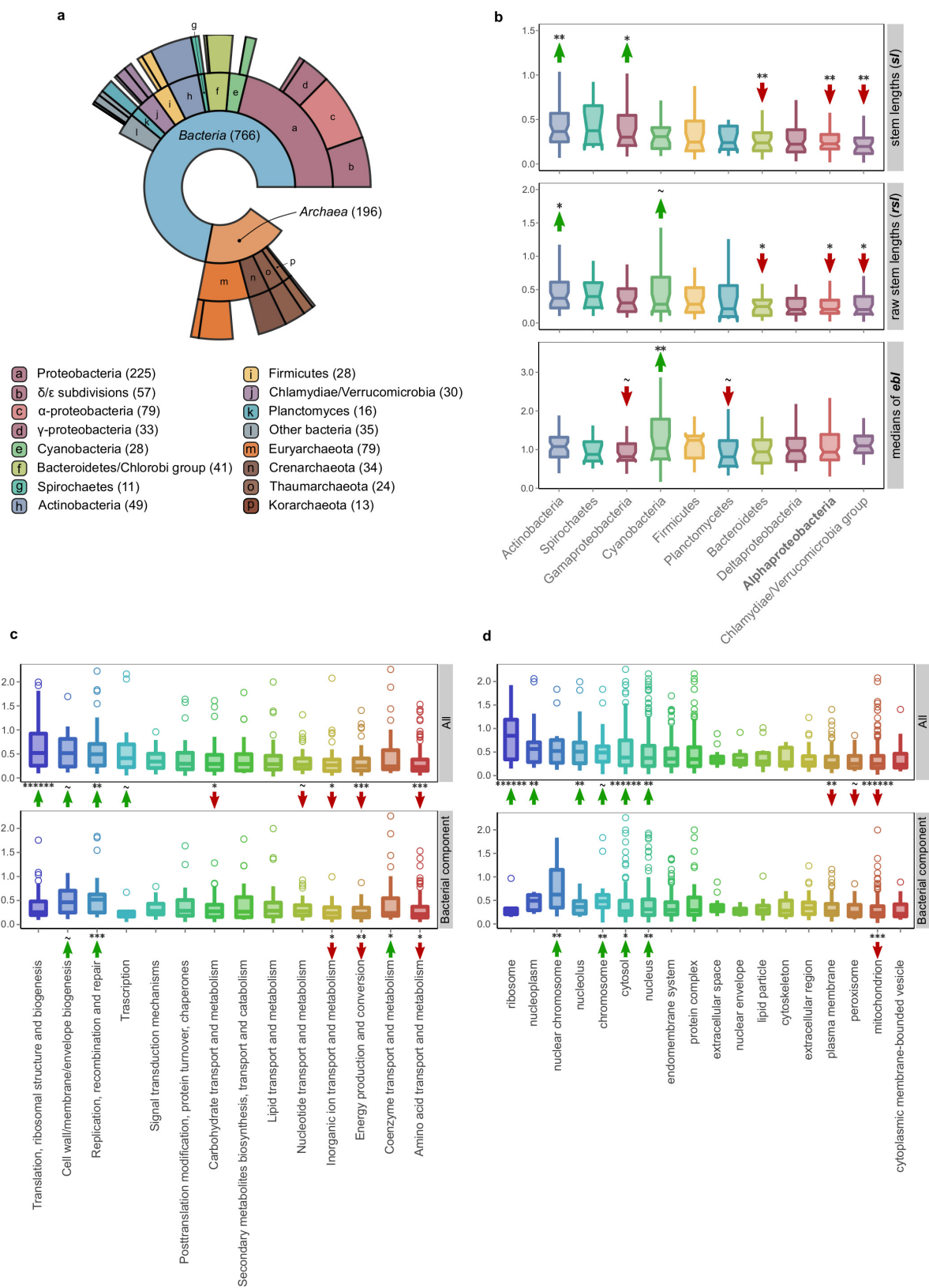
could bias the statistical inference, we used a stricter approach, considering only GO slim terms that were assigned to sequences from more than one group among unikonts, Archaeplastida, Chromalveolates and Excavates. Finally, through the corresponding orthologous groups, COG functional categories and GO slim terms were linked to prokaryotic groups and stem lengths, which were later used for profile comparison (Fig. 2c, d and Extended Data Fig. 1c, d). For convenience, we list here the COG functional categories corresponding to the one-letter codes: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control and mitosis; E, amino-acid metabolism and transport; F, nucleotide metabolism and transport; G, carbohydrate metabolism and transport; H, coenzyme metabolism; I, lipid metabolism; J, translation; K, transcription; L, replication and repair; M, cell wall/membrane/envelop biogenesis; N, cell motility; O, post-translational modification, protein turnover, chaperone functions; P, inorganic ion transport and metabolism; Q, secondary structure; T, signal transduction; U, intracellular trafficking and secretion; Y, nuclear structure; Z, cytoskeleton; R, general functional prediction only; S, function unknown.

**Unsupervised clustering and enrichment analyses.** The clustering of the stem lengths into different components was done by fitting a Gaussian mixture model using the expectation-maximization algorithm as implemented in the Mclust package<sup>29</sup> in R. The Mclust function returns the optimal model—the optimal number of components and membership—according to a maximum likelihood estimation and the Bayesian information criterion for expectation-maximization, initialized by hierarchical clustering for parameterized Gaussian mixture models. Applying the algorithm to the distribution of the normalized stem lengths from the LECA inference clustered the data into five components/subpopulations, of which the fifth, with only 14 extreme observations (with values in the range 2.3–7.1), also enriched in archaeal origins, was considered an outlier and was excluded. Each of the 1,064 remaining LECA families was assigned a membership within the four remaining components. Each of these subgroups was tested for enrichment in prokaryotic groups of origin, COG functional categories and GO cellular component terms. Enrichments were calculated using  $10^6$  permutations, in which the family memberships were randomly reshuffled and the  $P$  values estimated as the number of times a given origin, COG category or GO term had a count in the given component equal or greater than the observed one (Table 1).

**Statistical comparisons of stem lengths.** The statistical significance of the observed differences in normalized stem lengths between the different groups (taxonomic groups or functional categories and GO terms) was assessed with a non-parametric two-sided Mann–Whitney  $U$ -test for pairwise, or among three, comparisons. In the case of comparisons among three groups, the  $P$  values were adjusted for multiple testing with a correction for false discovery rate using the p.adjust function in R. The significance of the observed difference between the normalized stem lengths associated with the various groups and the overall bacterial signal was assessed using a permutation test with  $10^6$  randomizations. In each round, by sampling the whole distribution, the values were randomly assigned to the various eukaryotic families, and the mean, resulting from the random sampling of each of the groups, was computed (every group in each round had the same size but random values). The  $P$  value for each group was calculated as the number of times that an equal or more-extreme mean than the observed was occurring by chance, divided by the overall number of randomizations.

**Statistical associations.** We used a permutation test ( $10^6$  permutations) to evaluate the relationships between the proteins' evolutionary origin and their function/subcellular localization. The observed association was estimated as the number of co-occurrences between a given term and a given prokaryotic group of origin throughout all the families. The  $P$  value was calculated as the number of times that the amount of random co-occurrences between a group-term pair was equal or higher than the observed, divided by the number of permutations. Correspondence analysis is a statistical multivariate technique, conceptually similar to principal component analysis, that has been widely used to visualize associations between categorical variables<sup>30</sup>. Briefly, it decomposes the  $\chi^2$  statistic associated with the two-way table into orthogonal factors that maximize the separation between row and column scores. Correspondence analysis was applied to the contingency table of co-occurrences between the inferred taxonomic groups of prokaryotic origins (rows) and the various annotation terms (columns). The biplots in Fig. 3 and Extended Data Fig. 8a, b show the best two-dimensional approximation (first two principal axes) of the distances between rows and columns in each case. For the computation we used the ca function of the ca package in R, after removing very rare observations (single observation columns) that could bias the representation.

- Powell, S. *et al.* eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* **42**, D231–D239 (2014).
- Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
- Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
- Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
- Gabaldón, T. & Koonin, E. V. Functional and evolutionary implications of gene orthology. *Nature Rev. Genet.* **14**, 360–366 (2013).
- Huerta-Cepas, J. *et al.* PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res.* **39**, D556–D560 (2011).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Subramanian, A. R., Kaufmann, M. & Morgenstern, B. DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. *Algorithms Mol. Biol.* **3**, 6 (2008).
- Wallace, I. M., O'Sullivan, O., Higgins, D. G. & Notredame, C. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* **34**, 1692–1699 (2006).
- Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
- Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
- Huerta-Cepas, J., Dopazo, J. & Gabaldón, T. ETE: a python Environment for Tree Exploration. *BMC Bioinformatics* **11**, 24 (2010).
- Keeling, P. J. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant Biol.* **64**, 583–607 (2013).
- Fräley, C., Raftery, A. E., Murphy, T. B. & Scrucca, L. *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*. Technical Report No. 597 (Department of Statistics, Univ. Washington, 2012).
- Greenacre, M. *Correspondence Analysis in Practice* (Chapman & Hall, 2007).

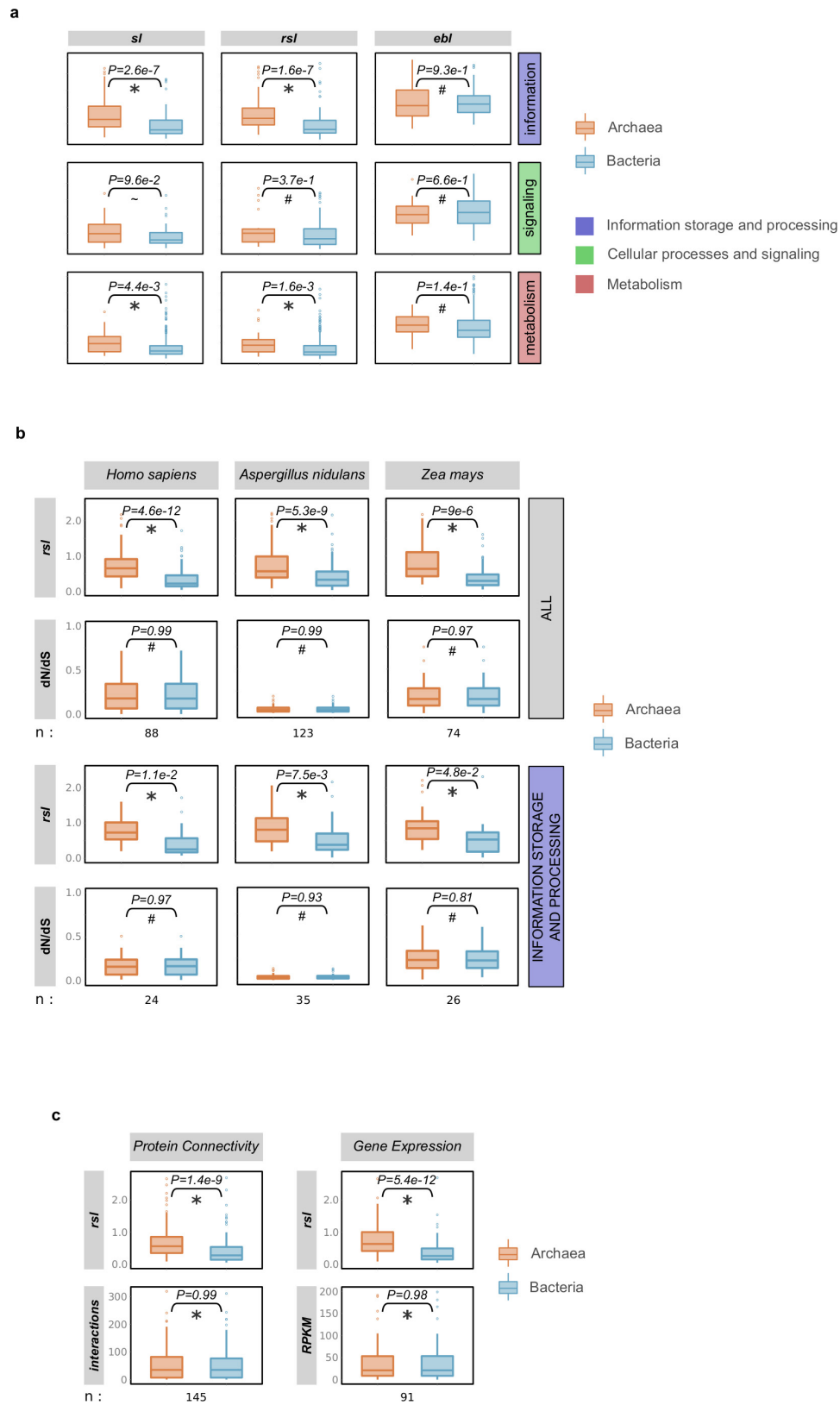


Extended Data Figure 1 | See next page for figure caption.



**Extended Data Figure 1 | Sister group distribution and extended phylogenetic distance profiles.** **a**, Ring plot showing the distribution of inferred prokaryotic origins. Inner layers represent hierarchically lower (broader) taxonomic levels. The number of LECA families assigned to each group is indicated in parentheses next to the corresponding level in the ring plot or in the boxes below. **b**, Box plot showing the distributions of branch lengths in the different bacterial components. Measured stem lengths (sl), raw stem lengths (rsl), and the medians of the lengths from LECA to branch tips inside the eukaryotic families (ebl), as defined in Fig. 1a, are shown. Permutation tests were performed to evaluate the statistical significance of the differences between the distributions. A total of  $10^6$  permutations were performed, with the values being randomly shuffled in each permutation (see also Methods). The arrows

and symbols above the boxes refer to the statistical significance of the differences observed compared with randomly shuffled distributions (lower values, downward red arrow; higher values, upward green arrow). The correspondence between the symbols and the  $P$  values is as follows:  $\sim P \leq 1 \times 10^{-1}$ ;  $*P \leq 5 \times 10^{-2}$ ;  $**P \leq 1 \times 10^{-2}$ ;  $***P \leq 1 \times 10^{-3}$ ;  $*****P < 1 \times 10^{-6}$ . The lower and upper box limits correspond to the first and third quartiles (the 25th and 75th percentiles). **c**, **d**, Stem length profiles of the various functional categories (**c**) and GO slim cellular components (**d**) are shown. As in Fig. 2c, the stem lengths are also evaluated by looking only at the bacterial component to exclude the possibility that the observed differences are due solely to archaeal–bacterial differences. The significance was assessed with permutation tests ( $10^6$  permutations) and is indicated with arrows as in **b**.

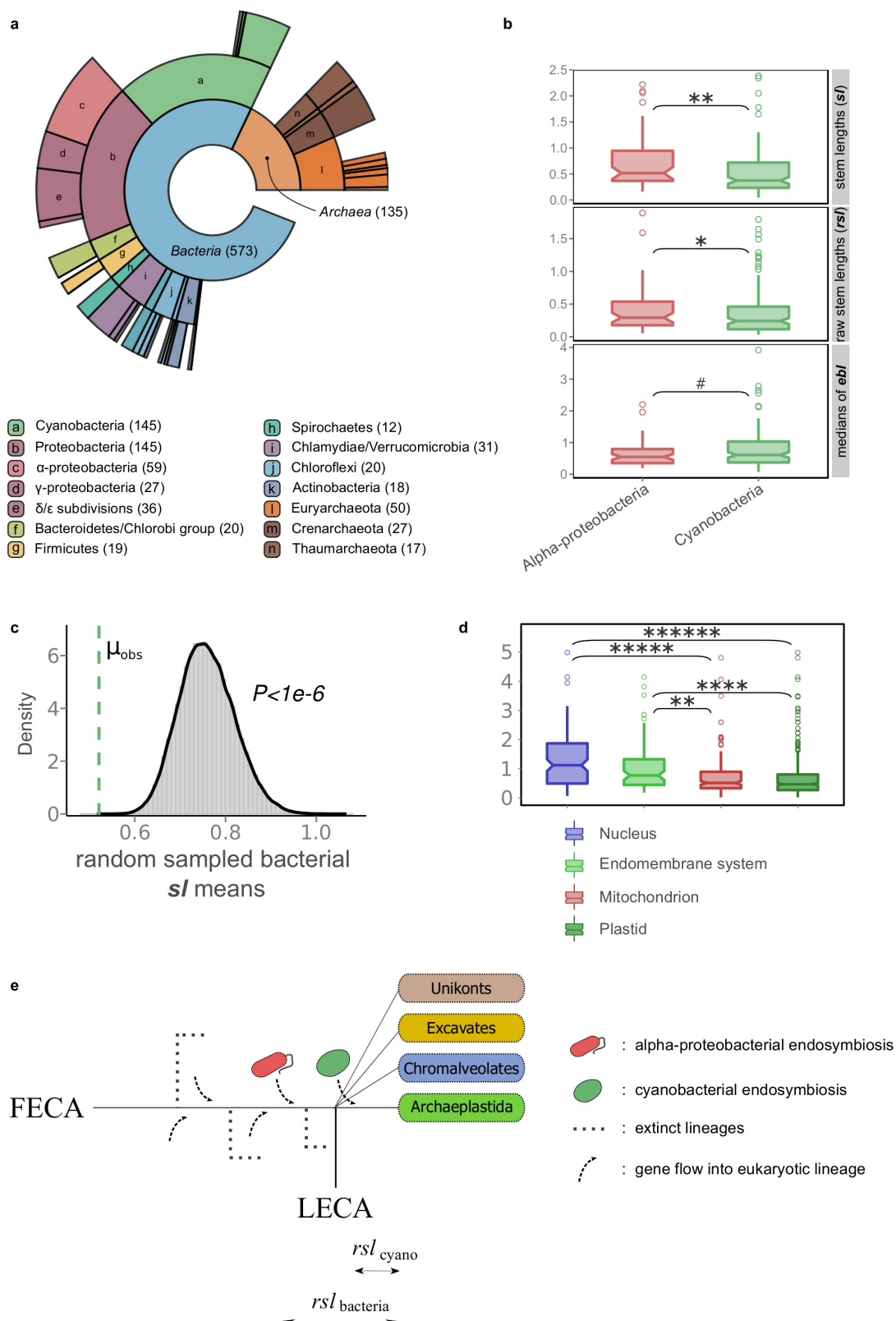


Extended Data Figure 2 | See next page for figure caption.

**Extended Data Figure 2 | Families of archaeal origin have significantly longer stems than families of bacterial origin across different functional categories, similar selective pressures, and connectivities/ expression levels. a,** The stem lengths, raw stem lengths, and eukaryotic branch lengths, between families of archaeal and bacterial inferred origin, are compared across the three major functional categories. While the eukaryotic branch lengths among the groups do not show significant differences, differences are detected in their respective stems (raw stem lengths and stem lengths). **b,** Archaeal and bacterial LECA families of similar selective pressures (as measured by dN/dS values across family members) differ significantly in terms of their raw stem lengths. Sets of families from both groups were matched with respect to their dN/dS values in the indicated reference species. The dN/dS data were downloaded from

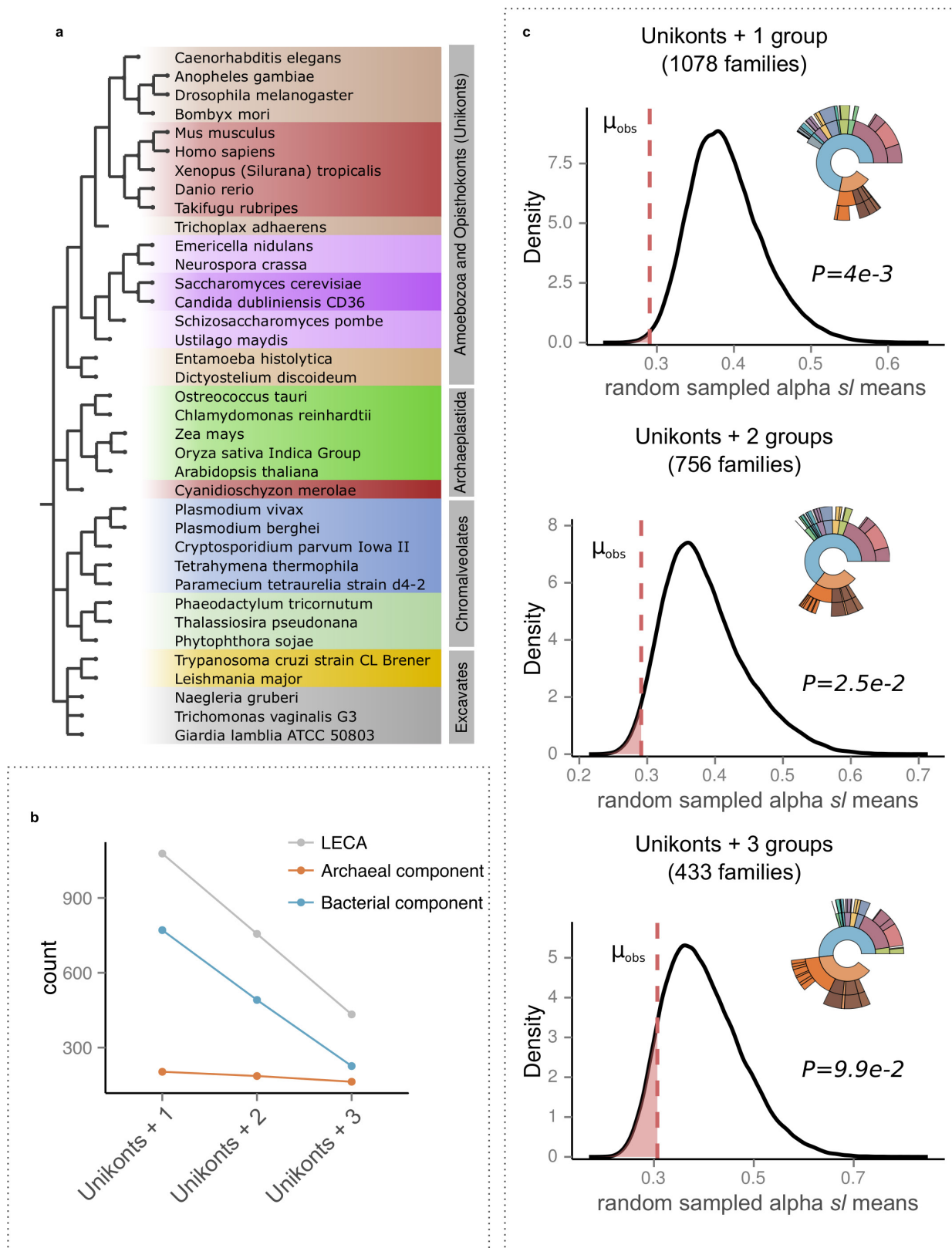
Ensembl for family members corresponding to *Homo sapiens* (Metazoa), *Aspergillus nidulans* (fungi) and *Zea mays* (plants) (see Supplementary Information section 1). The comparison of the raw stem lengths of the two sets shows that archaeal families generally have significantly longer stems (upper plots), and functions within the ‘information storage and processing’ category (lower plots), irrespective of their selective pressures. **c,** Archaeal and bacterial LECA families of similar connectivity/ expression levels show significantly different raw stem lengths (see Supplementary Information section 1). In **a–c**, differences between the archaeal and bacterial component were evaluated with a two-tailed Mann–Whitney *U*-test and the *P* value is indicated in each case (\* $P \leq 5 \times 10^{-2}$ ;  $\sim P \leq 1 \times 10^{-1}$ ; # $P > 1$ ).





**Extended Data Figure 3 | Analysis of the cyanobacterial signal in primary plastid-bearing eukaryotes.** **a**, Ring plot showing the distribution of inferred prokaryotic origins in widespread plant protein families, as in Extended Data Fig. 1a. The profile of inferred origins of eukaryotes that acquired a plastid through primary endosymbiosis carries a strong signal from the cyanobacterial endosymbiont. **b**, **c**, Families of inferred cyanobacterial origin have significantly shorter stem lengths and raw stem lengths than alphaproteobacterial families (**b**) and than the random distribution of stem lengths from the bacteria inferred component (**c**), pointing to a more recent acquisition of plastids

(post-LECA). **d**, Overall, as with mitochondrial localized proteins, those proteins localized to plastids have shorter stems than the nuclear and endomembrane system proteins. **e**, Schematic representation of the expected difference in stems, given that cyanobacterial endosymbiosis occurred after the diversification of the major eukaryotic lineages. As confirmed, the raw stem lengths measured from plant protein families to their common ancestor with cyanobacteria are shorter than those whose origin can be traced back to Alphaproteobacteria or other bacterial groups. Two-tailed Mann–Whitney *U*-test *P* value symbols in **b** and **d** are as in Extended Data Fig. 1; additionally \*\*\*\* $P \leq 1 \times 10^{-4}$ , \*\*\*\*\* $P \leq 1 \times 10^{-5}$ .

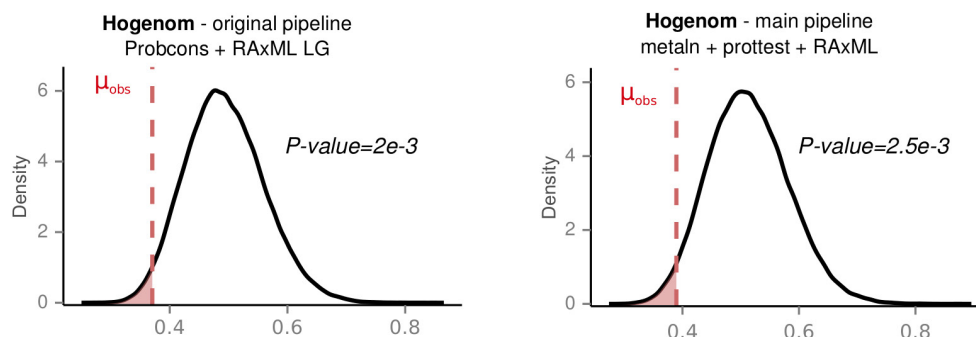
**Extended Data Figure 4 | Effect of alternative LECA definitions.**

**a**, The four eukaryotic groups including all 37 selected eukaryotic species used in the analysis are shown next to the NCBI taxonomic structure, with the higher groupings modified according to the Tree of Life Project (<http://tolweb.org/Eukaryotes/3>). **b**, Stricter LECA definitions have a much larger effect on the bacterial component than on the archaeal component, which is more widespread among eukaryotic groups. **c**, The effect of different LECA definitions in terms of taxonomic assignments

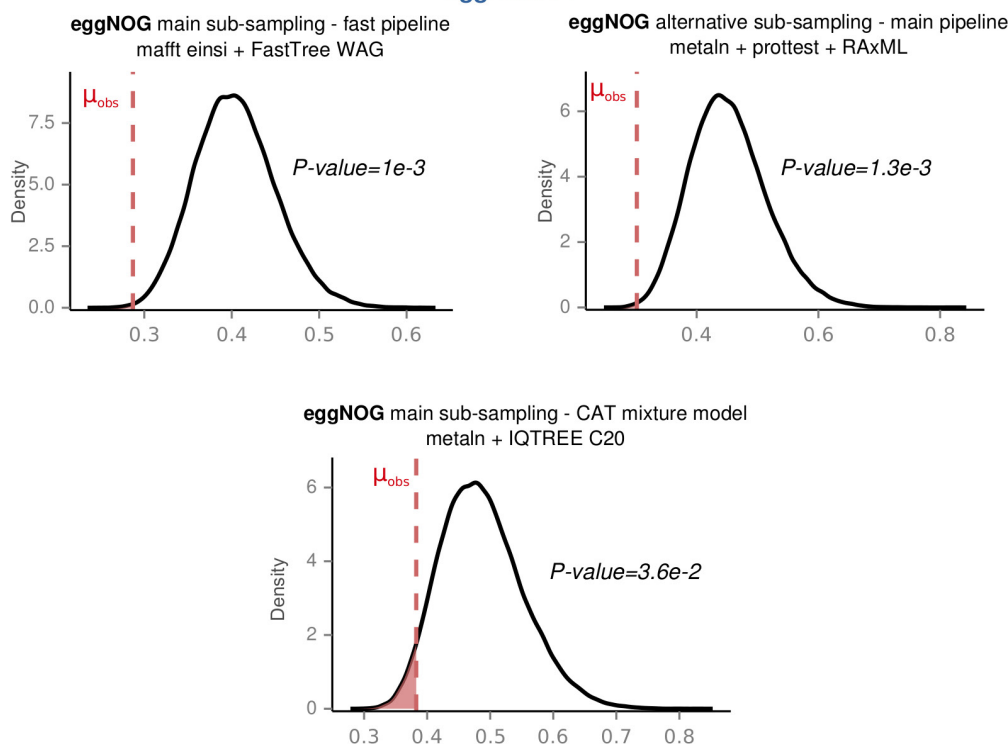
and differences in stem lengths between proteins of alphaproteobacterial origins and those derived from other bacteria. Numbers in parenthesis indicate the total number of LECA families that passed the threshold. The kernel density plots, as in Fig. 2b, show the observed stem length means for Alphaproteobacteria compared with  $10^6$  random samplings among values in protein families of bacterial origin. The observed means ( $\mu_{obs}$ ) are shown with a dashed red line, reflecting the  $P$  value of each test, and indicated next to the plot. See also Supplementary Information section 3.1.

a

## HOGENOM 5 (Rochette et al.)



## eggNOG v4



b

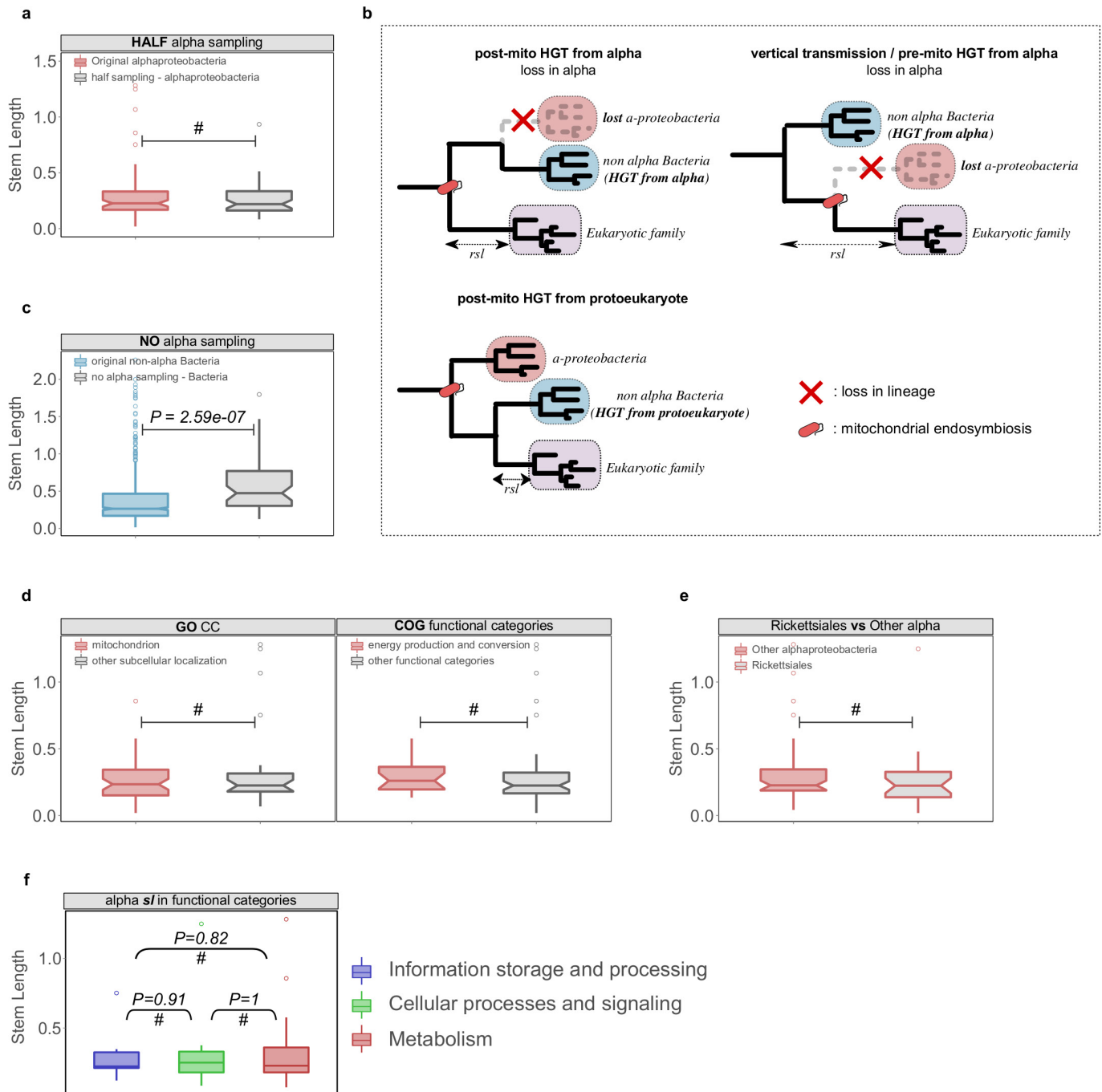
Support threshold	Bacterial families	$\alpha$ -proteobacterial families	P value*
0.5	766	79	4.0e-3
0.6	724	71	1.3e-3
0.7	669	64	5.9e-3
0.8	557	50	9.9e-3
0.9	354	31	7.0e-2

\*  $10^6$  permutations

**Extended Data Figure 5 | Alphaproteobacterial-derived proteins have consistently shorter branches, irrespective of the methods, data sets, and support thresholds.** Kernel density plots of the random mean distributions of the stem lengths are shown for the different methods, data sets and support thresholds used (see also Supplementary Information sections 3.2 and 3.3). The observed alphaproteobacterial means ( $\mu_{\text{obs}}$ ) are as in Fig. 2b. **a**, Results after using either the phylogenetic trees provided

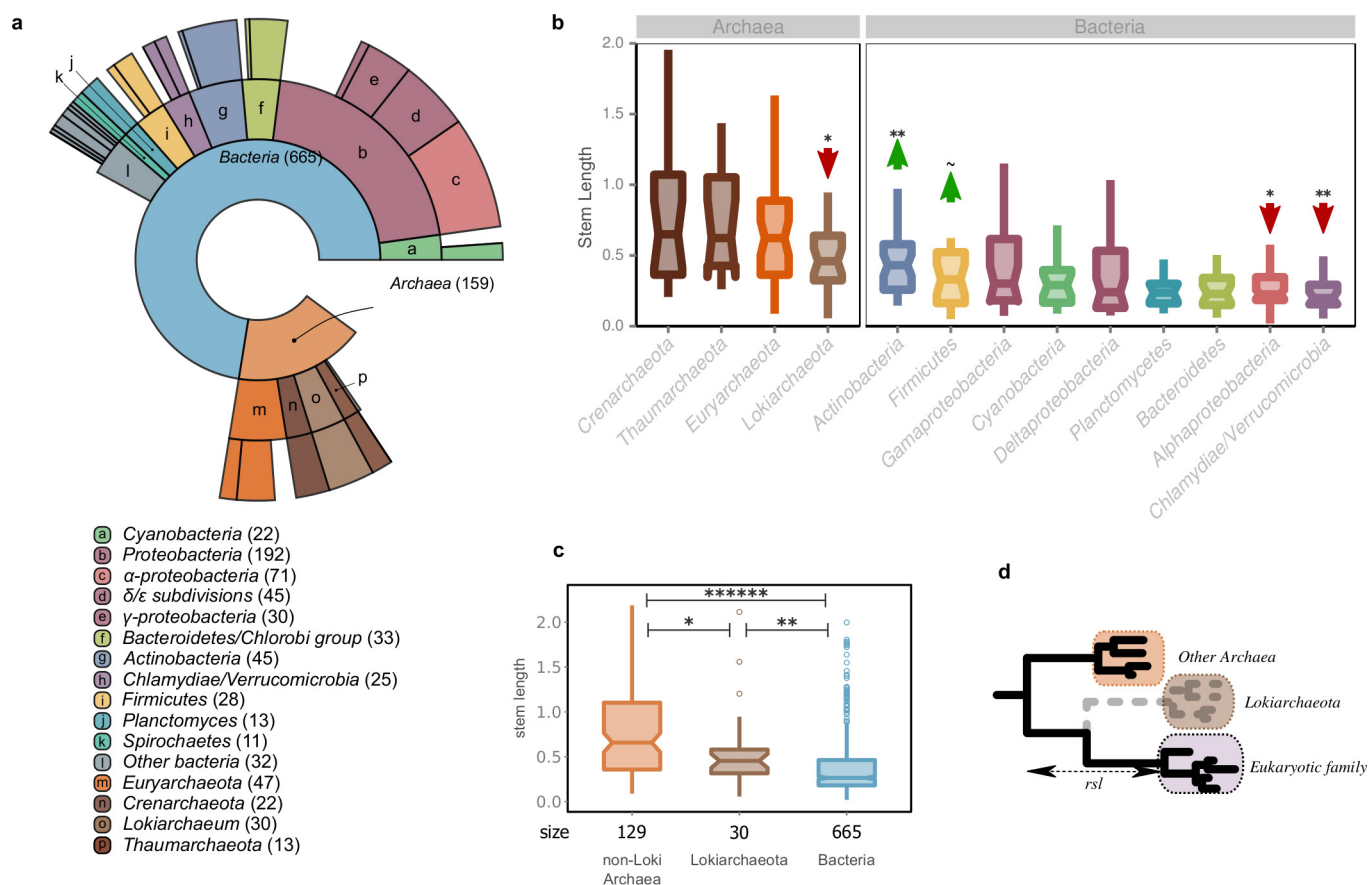
by the authors in ref. 8 (upper left), our standard phylogenetic pipeline applied to their sampling of sequences (upper right) or alternative phylogenetic pipelines or samplings from EggNOG (lower). **b**, The main result is robust against progressively stricter support thresholds until the sample size becomes too small (support threshold > 0.9). Numbers in parenthesis indicate the number of bacteria-inferred LECA families for each threshold.





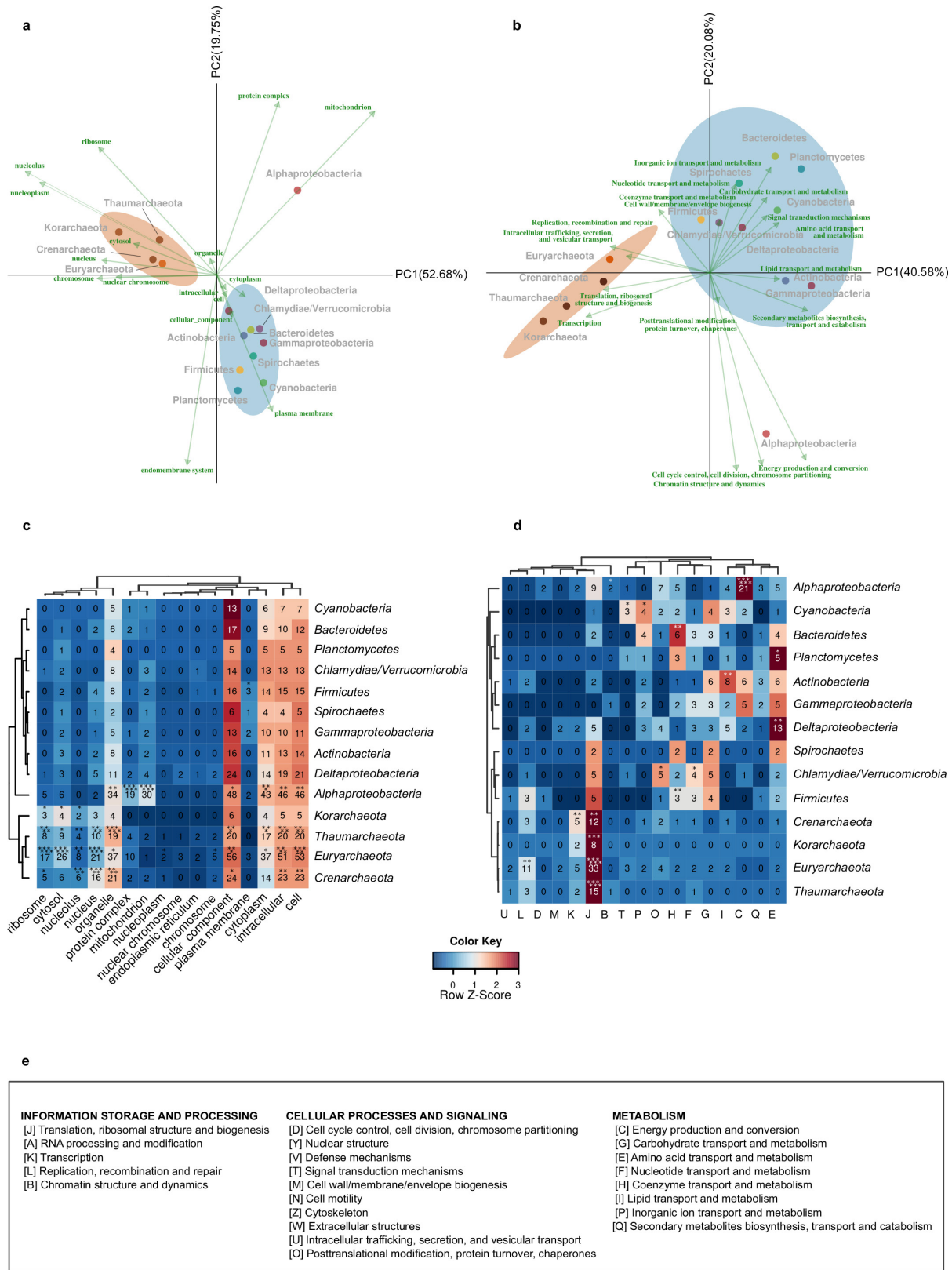
**Extended Data Figure 6 | Evaluation of alternative HGT scenarios and other potential biases.** **a**, The sampling effect was simulated by artificially removing part or all of the alphaproteobacterial sequences in the final data sets. To simulate the potential bias caused by an enriched sampling of Alphaproteobacteria, an artificial reduction of alphaproteobacterial sequences to 50% was applied to the data set ('HALF alpha sampling'). The reduction of alphaproteobacterial sequences by 50% does not significantly change the inferred stem length within families of alphaproteobacterial origin. #Cases where the difference was not significant. **b**, Different scenarios of HGT to the proto-mitochondrion are unable to explain the observed signal in families mapped to non-alpha Bacteria. The transfer of a gene from Alphaproteobacteria to another bacterial lineage after mitochondrial endosymbiosis and its parallel loss from the lineage of the mitochondrial ancestor ('post-mito HGT from alpha') would result in unchanged stem lengths. Loss of a gene from the alphaproteobacterial sister clade would result in an increase of the inferred stem lengths ('vertical transmission/pre-mito HGT from alpha'). The transfer of a gene from the protoeukaryotic lineage to other bacterial clades would result in shorter stem lengths compared with the

alphaproteobacterial mappings ('post-mito HGT from protoeukaryote'). **c**, Upon total exclusion of alphaproteobacterial sequences ('NO alpha sampling'), eukaryotic families map to other bacterial groups but with stem length higher than those observed typically. The same is observed when comparing the stem lengths of the families mapping to proteobacterial groups in the absence of Alphaproteobacteria with those typically mapping to proteobacterial groups other than Alphaproteobacteria. **d**, Box plots showing that there are no significant differences in the stem lengths between alphaproteobacterial families with mitochondrial localization compared with those with other subcellular localizations (left), or between families involved in energy-related functions compared with those involved in other functional categories (right). **e**, Box plot showing no significant difference between the distribution of stem lengths of families of Rickettsiales-inferred origin and other Alphaproteobacteria. **f**, Alphaproteobacterial families in different functional categories show no difference in stem lengths. In all cases the distributions were compared using a two-sided Mann-Whitney *U*-test. See also Supplementary Information sections 4 and 5.



**Extended Data Figure 7 | LECA inference and Lokiarchaeota.** Results after the inclusion of Lokiarchaeota in our analysis. **a**, The distribution of the sister group inference among prokaryotic taxonomy is shown in a ring plot together with the number of families in each group in parentheses (as in Extended Data Fig. 1). **b**, Box plot showing the stem length profiles of the various prokaryotic groups. Lokiarchaeota show the lowest values among all archaeal groups but higher values than any bacterial group. The symbols correspond to the same *P* values explained in Extended Data Fig. 1 after applying a permutation test ( $10^6$  permutations) for the archaeal and bacterial components, independently. **c**, Box plot with the comparison between the non-Loki archaeal, the Lokiarchaeota and the bacterial

stem length profiles. The *P* value symbols are as before (two-sided Mann–Whitney *U*-test, correction for false discovery rate). **d**, Schematic representation of the effect of the absence of *Lokiarchaeum* sequences on the stem lengths. The inferred origin of 30 eukaryotic families that were previously mapped to other, mainly archaeal, groups within the eggNOG version 4 database, is Lokiarchaeota, when homologous sequences from this metagenome are included. A reduction in the observed stem lengths of the families of Lokiarchaeota-inferred origin is expected in the scenario of Lokiarchaeota being the closest known archaeal relative of Eukaryotes. See also Supplementary Information section 6.



**Extended Data Figure 8 | Correspondence of different LECA components with different cellular localizations and functions (extended version of Fig. 3).** **a–d**, Different LECA components have different GO cellular components (**a, c**) and functional (**b, d**) profiles. Genes of different origin tend to have different functions and subcellular localizations. **a, b**, The same correspondence analysis symmetrical biplots as in Fig. 3 in higher resolution, with the names of the taxonomic group, the function and the GO slim terms indicated next to the coordinates. The percentage of variance explained by each principal component is indicated next to each axis in

parentheses. **c, d**, The contingency tables also used in correspondence analysis are shown in the form of a heatmap. The asterisks in the different cells reflect the significance of the association between a given origin and a localization (**c**) or function (**d**), as computed using permutation tests ( $10^6$  permutations), where the annotations among each eukaryotic family were reshuffled (see Methods). The correspondence between the symbols and the *P* values is as in Extended Data Figs 1 and 3. **e**, The COG functional categories, as organized in the three major groups ‘information storage and processing’, ‘cellular processes and signalling’ and ‘metabolism’.



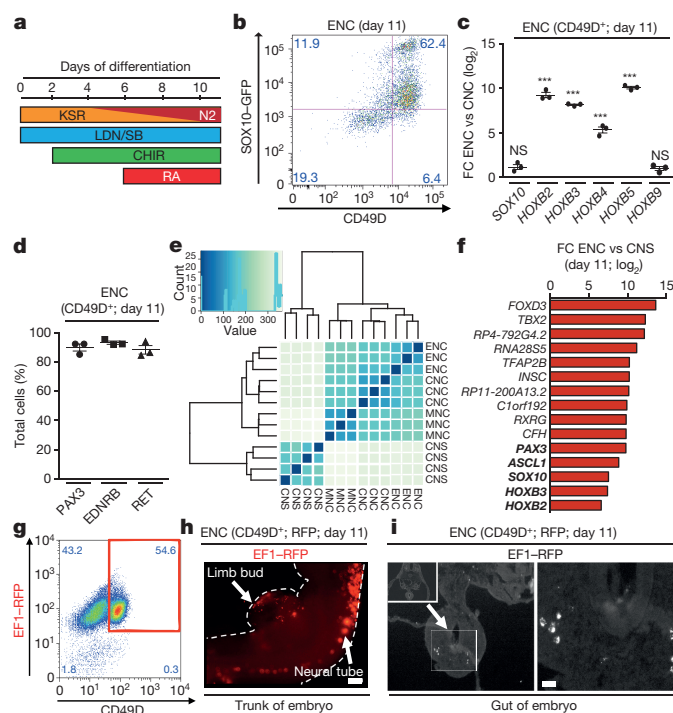
# Deriving human ENS lineages for cell therapy and drug discovery in Hirschsprung disease

Faranak Fattahi<sup>1,2,3</sup>, Julius A Steinbeck<sup>1,2</sup>, Sonja Kriks<sup>1,2</sup>, Jason Tchieu<sup>1,2</sup>, Bastian Zimmer<sup>1,2</sup>, Sarah Kishinevsky<sup>1,2,3</sup>, Nadja Zeltner<sup>1,2</sup>, Yvonne Mica<sup>1,2†</sup>, Wael El-Nachef<sup>4</sup>, Huiyong Zhao<sup>4</sup>, Elisa de Stanchina<sup>4</sup>, Michael D. Gershon<sup>5</sup>, Tracy C. Grikscheit<sup>6</sup>, Shuibing Chen<sup>7</sup> & Lorenz Studer<sup>1,2</sup>

The enteric nervous system (ENS) is the largest component of the autonomic nervous system, with neuron numbers surpassing those present in the spinal cord<sup>1</sup>. The ENS has been called the ‘second brain’<sup>1</sup> given its autonomy, remarkable neurotransmitter diversity and complex cytoarchitecture. Defects in ENS development are responsible for many human disorders including Hirschsprung disease (HSCR). HSCR is caused by the developmental failure of ENS progenitors to migrate into the gastrointestinal tract, particularly the distal colon<sup>2</sup>. Human ENS development remains poorly understood owing to the lack of an easily accessible model system. Here we demonstrate the efficient derivation and isolation of ENS progenitors from human pluripotent stem (PS) cells, and their further differentiation into functional enteric neurons. ENS precursors derived *in vitro* are capable of targeted migration in the developing chick embryo and extensive colonization of the adult mouse colon. The *in vivo* engraftment and migration of human PS-cell-derived ENS precursors rescue disease-related mortality in HSCR mice (*Ednrb*<sup>-/-</sup>), although the mechanism of action remains unclear. Finally, EDNRB-null mutant ENS precursors enable modelling of HSCR-related migration defects, and the identification of pepstatin A as a candidate therapeutic target. Our study establishes the first, to our knowledge, human PS-cell-based platform for the study of human ENS development, and presents cell- and drug-based strategies for the treatment of HSCR.

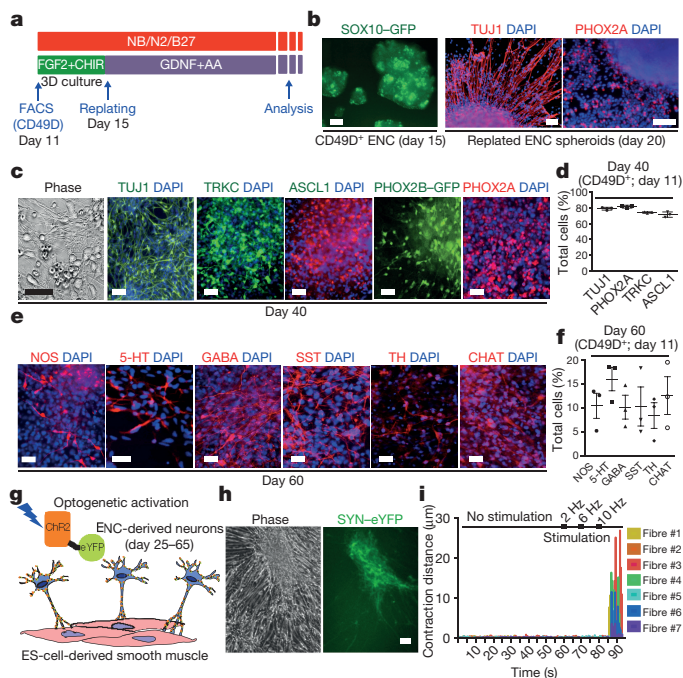
The *in vitro* derivation of human ENS lineages from human PS cells has remained elusive despite the great medical needs. The ENS develops from both the vagal and sacral neural crest (NC). We focused on the vagal NC that generates most of the ENS and migrates caudally to colonize the entire length of the bowel<sup>2</sup>. Current NC differentiation protocols<sup>3</sup> result in SOX10<sup>+</sup> NC precursors that are HOX negative (indicative of anterior/cranial identity; cranial neural crests (CNCs)). By contrast, vagal NC identity is characterized by the expression of specific, regionally restricted HOX genes including *Hoxb3* (ref. 4) and *Hoxb5* (ref. 5). Retinoic acid has been used previously as an extrinsic factor to shift the regional identity of central nervous system (CNS) precursors from anterior to more caudal fates during motor neuron specification<sup>6</sup>. Here we tested whether the addition of retinoic acid can similarly direct the regional identity in NC lineages and induce the expression of vagal markers (Fig. 1a). After treatment of SOX10–green fluorescent protein (SOX10–GFP) ES cells (ref. 3) with retinoic acid, we obtained GFP<sup>+</sup> cells with yields comparable to CNC conditions, indicating that retinoic acid treatment does not interfere with overall NC induction (Fig. 1b and Extended Data Fig. 1a, b). To facilitate the isolation of pure NC populations, we performed a candidate surface marker screen and identified CD49D ( $\alpha 4$  integrin) as an epitope that reliably marks early SOX10<sup>+</sup> NC lineages (Fig. 1b and Extended Data Fig. 1a–c). We next used CD49D

to demonstrate the robustness of retinoic-acid-based NC induction across human PS cell lines (both human embryonic stem (ES) cells



**Figure 1 | Deriving ENS precursors from human ES cells.** **a**, Protocol (days 0–11) for deriving enteric NC (ENC) cells. CHIR, CHIR99021; KSR, knockout serum replacement; LDN/SB, medium containing LDN193189 and SB431542; RA, retinoic acid. **b**, Flow cytometry of ENC for SOX10–GFP and CD49D at day 11. **c**, Quantitative reverse transcriptase PCR (qRT–PCR) for SOX10, vagal NC markers *HOXB2–5* and *HOXB9* in CD49D<sup>+</sup> ENC versus CNC cells;  $n = 3$  independent experiments. FC, fold change. **d**, Quantification of PAX3, RET and EDNRB immunofluorescence in CD49D<sup>+</sup> ENC lineage;  $n = 3$  independent experiments. **e**, Unsupervised clustering of CD49D<sup>+</sup> NC versus matched CNS precursor (day 11). **f**, Top 10 and selected additional (bold) differentially expressed transcripts in CD49D<sup>+</sup> ENC versus CNS precursors. *C1orf192* is also known as *CFAP126*; *RP4-792G4.2* is an antisense transcript to *FOXD3*; *RP11-200A13.2* is a pseudogene. **g**, RFP<sup>+</sup> and CD49D<sup>+</sup> ENC precursors are FACS-purified (day 11) for transplantation into developing chick embryos. **h**, Whole-mount epifluorescence showing distribution of RFP<sup>+</sup> cells 24 h after injection. **i**, Cross section of the embryos at trunk levels shows RFP<sup>+</sup> cells located in the gut anlage (left) and at higher magnification (right). Scale bars, 200  $\mu$ m (**h**) and 10  $\mu$ m (**i**). Data are mean  $\pm$  s.e.m. NS, not significant. \*\*\* $P < 0.001$  ( $t$ -test, ENC compared to CNC;  $n = 3$ ).

<sup>1</sup>The Center for Stem Cell Biology, New York, New York 10065, USA. <sup>2</sup>Developmental Biology Program, Sloan-Kettering Institute for Cancer Research, New York, New York 10065, USA. <sup>3</sup>Weill Graduate School of Medical Sciences of Cornell University, New York, New York 10065, USA. <sup>4</sup>Molecular Pharmacology Program, New York, New York 10065, USA. <sup>5</sup>Department of Pathology and Cell Biology, Columbia University, College of Physicians and Surgeons, New York, New York 10032, USA. <sup>6</sup>Children's Hospital Los Angeles, Pediatric Surgery, Los Angeles, California 90027, USA. <sup>7</sup>Department of Surgery, Weill Medical College of Cornell University, New York, New York 10065, USA. <sup>†</sup>Present address: Thermo Fisher Scientific, Waltham, Massachusetts 02451, USA.



**Figure 2 | Differentiation of human ES-cell-derived ENC precursors into enteric neuron subtypes.** **a**, Protocol for neuronal differentiation and maturation of ENC precursors. AA, ascorbic acid; NB/N2/B27, neurobasal medium with N2 and B27 supplement. **b**, SOX10-GFP-expressing 3D spheroids from purified ENC precursors gave rise to TUJ1 and PHOX2A enteric neuron lineage. **c**, **d**, Phase-contrast and immunofluorescence images (c) and quantification (d) at day 40. ENC-derived cells express TRKC, ASCL1 and PHOX2A/B. PHOX2B expression was confirmed using H9 human ES-cell-based PHOX2B-GFP reporter line;  $n = 3$  independent experiments. **e**, **f**, Immunofluorescence analysis (e) and quantification (f) for expression of diverse neurotransmitters. Cells were derived from FACS-purified, CD49D<sup>+</sup> ENC precursors to ensure NC origin;  $n = 3$  independent experiments. SST, somatostatin; TH, tyrosine hydroxylase. **g**, Light-stimulated activation of ENC-derived neurons expressing channelrhodopsin-2 (ChR2). **h**, Phase-contrast and live fluorescence images of human ES-cell-derived smooth muscle cells (SMCs) and ENC-derived neuron co-cultures subjected to light stimulation. **i**, Diagram representing extent of contraction of SMCs before and during light (450 nm) stimulation at increasing frequencies. Scale bars, 100  $\mu$ m (b) and 50  $\mu$ m (c, e, h). Data are mean  $\pm$  s.e.m.

and induced pluripotent stem (iPS) cells; Extended Data Fig. 1d). Purified CD49D<sup>+</sup> NC precursors, derived in the presence of retinoic acid, expressed *HOXB2–HOXB5* indicative of vagal identity<sup>4,5</sup>, but not more caudal HOX transcripts such as *HOXB9* (Fig. 1c). In further agreement with vagal identity, CD49D<sup>+</sup>, retinoic-acid-treated NC precursors expressed markers of early enteric NC (ENC) lineages<sup>2</sup> including PAX3, EDNRB and RET (Fig. 1d and Extended Data Fig. 1e, f). Given the paucity of developmental data on human ENC development, we performed RNA sequencing (RNA-seq) analysis in ES-cell-derived ENC precursors, in CNCs (no retinoic acid), in melanocyte-biased<sup>3</sup> NCs (MNCs) (Extended Data Fig. 1a), and in stage-matched CNS precursors<sup>7</sup>. Unsupervised clustering reliably segregated the transcriptomes of all PS-cell-derived NC populations away from CNS precursors and further subdivided the various NC sublineages (Fig. 1e). The most differentially expressed genes in the ENC compared to CNS lineage included general NC markers such as *FOXD3* or *TFAP2A* but also *PAX3* and *HOX* genes related to the ENC lineage (Fig. 1f). CNCs and MNCs were also enriched in general NC markers but showed high levels of *NEUROG1*, *ISL1* or *MLANA*, *TYR* and *DCT* expression respectively, compatible with their subtype identity (Extended Data Fig. 1g, h). Direct comparison of the various NC lineages yielded novel candidate marker of human vagal NC/ENC lineage (Fig. 1f). A list

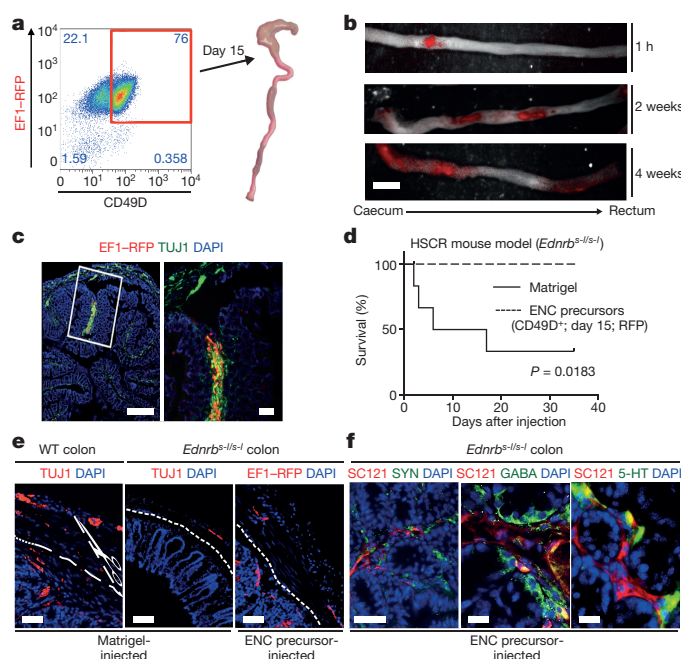
of the top 200 enriched transcripts for each NC lineage is provided (Supplementary Tables 1–3).

One key functional property of the ENC is the ability to migrate extensively and to colonize the gut<sup>2</sup>. Red fluorescent protein (RFP)-labelled, CD49D<sup>+</sup>-purified (Fig. 1g) PS-cell-derived ENC precursors were injected into the developing chick embryo at the level of the vagal NC. Transplanted human cells migrated along the trunk of the embryo (Fig. 1h) and colonized the gut (22 out of 57 embryos injected; Fig. 1i). By contrast, stage-matched CNC or MNC precursors targeted cranial regions (CNC) or followed a trajectory along the dermis (MNC) (Extended Data Fig. 1i).

To address whether ES-cell-derived ENC precursors are capable of recreating ENS neuronal diversity, we maintained purified CD49D<sup>+</sup> ENC precursors in 3D spheroids for 4 days followed by differentiation as adherent cultures in the presence of ascorbic acid and glial-cell-line-derived neurotrophic factor (GDNF) (Fig. 2a). The 3D spheroid step was required to retain high levels of SOX10-GFP expression (Fig. 2b). Replating of 3D spheroids under differentiation conditions yielded immature neurons expressing TUJ1 and the enteric precursor marker PHOX2A (day 20; Fig. 2b). Most PHOX2A<sup>+</sup> cells were positive for TRKC (also known as NTRK3), a surface marker expressed in enteric neuron precursors<sup>8</sup> suitable for enrichment for PHOX2A<sup>+</sup> and ASCL1<sup>+</sup> precursors (Extended Data Fig. 2a, b). Temporal expression analyses (Extended Data Fig. 2c–e) showed maintenance of ENC neuronal precursor marker expression by day 40 of differentiation (Fig. 2c, d), followed by an increase in the percentage of mature neurons by day 60 (Fig. 2e, f). In agreement with enteric neuron identity, we observed a broad range of neurotransmitter phenotypes including serotonin-positive (5-hydroxytryptamine<sup>+</sup>, 5-HT<sup>+</sup>), GABA<sup>+</sup> ( $\gamma$ -aminobutyric-acid-positive) and nitric-oxide-synthase-positive (NOS)<sup>+</sup> neurons. The presence of these neurotransmitters in neurons derived from CD49D<sup>+</sup>-purified NC precursors indicates ENC origin, since those neurotransmitters are not expressed in other NC lineages. Indeed, no 5-HT<sup>+</sup> neurons were observed in parallel cultures derived from HOX-negative, CD49D<sup>+</sup> cells (Extended Data Fig. 3a, b). CNC-derived precursors differentiated into tyrosine-hydroxylase-expressing neurons (Extended Data Fig. 3c), and gave rise to TRKB-positive rather than TRKC-positive precursors, suggesting enrichment for sympathetic neuron lineages (Extended Data Fig. 3d, e).

A major function of the ENS is the control of peristaltic gut movements. We probed the functionality of *in-vitro*-derived enteric neurons by assessing connectivity with smooth muscle cells (SMCs). ES-cell-derived SMCs were generated via a mesoderm intermediate after exposure to activin A and BMP4 *in vitro*<sup>9</sup> and culture in the presence of TGF $\beta$  (Extended Data Fig. 4a). The resulting SMC progenitors expressed ISL1 and were immunoreactive for smooth muscle actin (SMA) (Extended Data Fig. 4b). For connectivity studies, an optogenetic reporter line (ref. 10) was used to allow for light-induced control of neuronal activity. Enteric neurons were derived from a human ES cell line expressing enhanced yellow fluorescent protein (eYFP)-tagged channelrhodopsin-2 (ChR2) under control of the human synapsin (SYN) promoter (Fig. 2g). GABA<sup>+</sup> and 5-HT<sup>+</sup> neurons in these co-cultures were closely associated with SMCs (Extended Data Fig. 4c). Interestingly, co-culture of day-25 neurons with SMCs triggered accelerated neuronal maturation, as illustrated by the increased expression of SYN-eYFP (Extended Data Fig. 4d). Conversely, ES-cell-derived SMCs also showed signs of accelerated maturation under co-culture conditions as illustrated by the expression of mature markers (MYH11 and acetylcholine receptor (AChR); Extended Data Fig. 4e) and the ability to contract in response to pharmacological stimulation (Supplementary Videos 1–6 and Extended Data Fig. 4f). While no spontaneous contractions were observed under co-culture conditions, a wave of SMC contractions could be triggered 5–10 s after light-mediated activation (10 Hz frequency) of SYN-ChR2-eYFP neurons (Fig. 2h, i and Supplementary Video 7). Notably, both light- and drug-induced SMC contractions





**Figure 3 | Human ENC precursors migrate extensively in normal and HSCR adult colon.** **a**, Transplantation model into the wall of proximal colon (caecum) of adult mice. **b**, Whole-mount fluorescence imaging of grafted RFP<sup>+</sup> human ES-cell-derived ENC precursors inside adult colon wall (NSG mice). **c**, Cross section of NSG colon showing robust survival and TUJ1 staining of grafted cells. **d**, Survival curve of *Ednrb*<sup>Δ/Δ</sup> (HSCR) animals grafted with human ES-cell-derived ENC precursors versus Matrigel-only. **e**, Immunofluorescence staining of normal or HSCR colons transplanted with Matrigel (control) or RFP<sup>+</sup> human ES-cell-derived ENC precursors showing distribution of human cells in myenteric and submucosal layers. Dashed lines indicate border between submucosal and muscle layers. WT, wild type. **f**, Representative images of grafted animals 3 months after transplantation into *Ednrb*<sup>Δ/Δ</sup> colon, showing expression of GABA and 5-HT in SC121 human cells. All *in vivo* experiments were performed in a blinded manner. *Ednrb*<sup>Δ/Δ</sup> mutation was confirmed in all grafted animals; *n* = 6 animals each for treatment and control groups. Scale bars, 1 cm (**b**), 500 μm (**c**, left); 100 μm (**c**, right, **e**) and 25 μm (**f**). *P* value for survival analysis is given numerically, log-rank (Mantel–Cox) test.

were slow and involved the movement of sheet-like structures, suggesting coordination among cells possibly via gap-junction-mediated coupling. These studies demonstrate functional connectivity between ES-cell-derived enteric neurons and SMCs. *In vivo* interactions of the ENS within the gut, however, are more complex and involve several cell types. As a first step in modelling those interactions in 3D, we used a tissue-engineering approach combining *in-vitro*-derived human ENC precursors (CD49D<sup>+</sup>, day 15) with mouse primary intestinal tissue (Extended Data Fig. 5a). Using our previously established protocols to form organoid units<sup>11</sup>, the recombined tissue constructs were seeded onto a scaffold and implanted onto the omentum of immunodeficient hosts for maturation *in vivo*<sup>11</sup>. Human cells were readily detected within gut-like structures using the human-specific markers SC121 and synaptophysin. Importantly, cells were located both in epithelial and muscle layers (Extended Data Fig. 5b), showing their ability to interact with both target cell types.

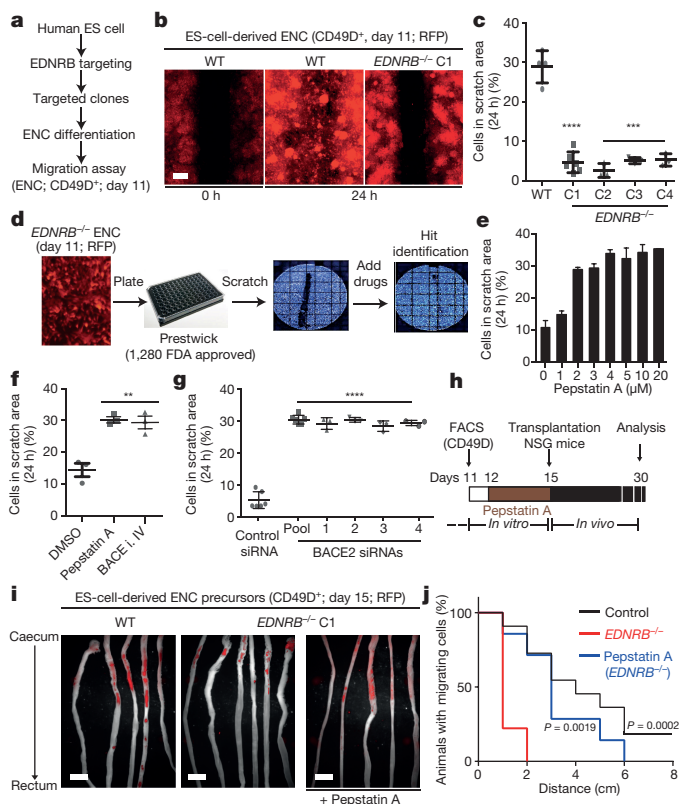
Transplantation of PS-cell-derived precursors could yield novel therapeutic opportunities for ENS disorders such as HSCR. Children with HSCR are currently treated by surgical removal of the aganglionic portion of the gut. While life-saving, the surgery does not address dysfunction of the remaining gastrointestinal tract in surviving patients<sup>12</sup>. Furthermore, therapeutic options in patients with total aganglionosis are limited. A major challenge in developing a cell therapy for HSCR is the need to repopulate the ENS over extensive distances. Previous studies tested the transplantation of a variety of candidate cell sources into the fetal or postnatal colon<sup>13,14</sup>. Mouse fetal-derived ENC precursors

resulted in the most promising data, with evidence of functional integration but limited *in vivo* migration<sup>15</sup>. To assess the ability of human ES-cell-derived ENC precursors to migrate within the postnatal or adult colon (3–6 weeks of age), we performed orthotopic injections of CD49D<sup>+</sup> RFP-labelled precursors into NOD-SCID-*Il2rg*<sup>−/−</sup> (NSG) mice (Fig. 3a). Cells were injected into the wall of the caecum aiming for the muscle layer (Fig. 3a) and resulting in a well-defined deposit of RFP<sup>+</sup> cells 1 h after injection (Extended Data Fig. 6a, left, and Fig. 3b, top). Notably, 2–4 weeks after transplantation, RFP<sup>+</sup> cells had migrated extensively and repopulated the host colon over its entire length (Fig. 3b). The grafted ENC precursors formed clusters along the colon (Extended Data Fig. 6a, right) expressing TUJ1 (Fig. 3c). By contrast, stage-matched CNS and CNC precursors, grafted under identical conditions, showed limited migration (Extended Data Fig. 6b, c).

Given the extraordinary ability of the grafted cells to repopulate the host colon, we tested whether those cells could provide therapeutic benefit in an HSCR model. One widely used genetic model is the *Ednrb*<sup>Δ/Δ</sup> mouse<sup>16</sup>. These mice develop a megacolon owing to aberrant peristalsis. As a consequence, mutant mice show high mortality by 4–6 weeks of age. *Ednrb*<sup>Δ/Δ</sup> mice were injected 2–3 weeks after birth with RFP<sup>+</sup> human ES-cell-derived ENC precursors (treatment group) or with Matrigel vehicle (control group). Most control-injected animals (*n* = 6) died over a period of 4–5 weeks (Fig. 3d) with a megacolon-like pathology (Extended Data Fig. 6d). By contrast, all animals injected with ES-cell-derived ENC precursors (*n* = 6) survived (Fig. 3d). Grafted animals were assessed for graft survival and migration at 6–8 weeks of age. Whole-mount fluorescence imaging confirmed the migration of ES-cell-derived ENC precursors within the HSCR colon (Extended Data Fig. 6e, f). Preliminary studies showed a trend towards an improved gastrointestinal transit time in grafted versus the small subset of Matrigel-treated animals that survived beyond 6 weeks of treatment, as measured using carmine dye gavage (Extended Data Fig. 6g). Histological analyses at 6 weeks and 3 months after transplantation confirmed myenteric and submucosal localization of human cells in HSCR colon. While the location mimicked aspects of endogenous ENS, there was a preference towards the submucosal region (Fig. 3e and Extended Data Fig. 6h, top). Human cells were also detected in the distal colon (Extended Data Fig. 6h, bottom), where only few endogenous TUJ1<sup>+</sup> cells were detected. Immunocytochemistry for SC121 and human-specific synaptophysin, not detected in Matrigel-injected animals (Extended Data Fig. 6i), confirmed human identity and pre-synaptic marker expression (Fig. 3f, right). In addition to neuronal cells (Extended Data Fig. 6h), we also observed human cells expressing glial markers such as GFAP (Extended Data Fig. 6j). Neuron-subtype-specific markers in SC121<sup>+</sup> human cells included 5-HT, GABA and NOS (Fig. 3f and Extended Data Fig. 6k, l).

Our findings demonstrate that wild-type human PS-cell-derived ENS precursors can repopulate the colon of HSCR mice. However, HSCR-causing mutations often affect the migration of ENS precursor in a cell autonomous manner<sup>2,17</sup>. Therefore, developing patient-matched cell therapies for HSCR may require complementary genetic or pharmacological strategies to overcome intrinsic migration defects of the transplanted cells. As causative genetic defects in HSCR patients are often not known or complex<sup>2,18</sup>, gene correction before transplantation may not be possible. We therefore assessed whether human PS-cell-derived ENS precursors can be used to model HSCR and serve as a platform to screen for candidate compounds that could overcome disease-related migration defects. In a first step, we established isogenic ES cell lines with homozygous loss-of-function mutations in *EDNRB* using CRISPR/Cas-based gene-targeting techniques<sup>19</sup> (Fig. 4a and Extended Data Fig. 7a, b). Loss-of-function mutations in *EDNRB* are a well-known genetic cause in a subset of HSCR patients<sup>20</sup>. ENC precursors could be derived at comparable efficiencies from *EDNRB*-mutant and control lines (Extended Data Fig. 7c). However, CD49D<sup>+</sup> ENC precursors from four *EDNRB*<sup>−/−</sup> clones showed a notable migration defect when using the scratch assay to model aspects of the HSCR phenotype<sup>21</sup>





**Figure 4 | EDNRB signalling regulates human ENC precursor cell migration.** **a**, *In vitro* HSCR disease modelling model. **b**, **c**, Representative images and quantification of scratch assay in RFP<sup>+</sup>/CD49D<sup>+</sup> wild-type and EDNRB<sup>-/-</sup> human ES-cell-derived ENCs (clones (C) 1–4);  $n = 3$  independent experiments. **d**, Illustration of chemical screen. **e**, Dose-response of pepstatin A on migration of CD49D<sup>+</sup> EDNRB<sup>-/-</sup> human ES-cell-derived ENCs. **f**, Quantification of CD49D<sup>+</sup> EDNRB<sup>-/-</sup> human ES-cell-derived ENC migration after treatment with pepstatin A (10 μM) or BACE inhibitor (i. IV) (1 μM);  $n = 3$  independent experiments. **g**, Quantification of cell migration after BACE2 knockdown using a pool of five different short interfering RNAs (siRNAs) or four individual siRNAs;  $n = 3$  independent experiments. **h**, Pepstatin A pre-transplantation treatment model. **i**, Whole-mount images of NSG colon transplanted with RFP<sup>+</sup> CD49D purified wild-type and EDNRB<sup>-/-</sup> human ES-cell-derived ENC precursors with or without pepstatin A pre-treatment. **j**, Quantification of the fraction of animals with human cells present in colon at increasing distance from injection site (see Extended Data Fig. 8c);  $n \geq 8$  animals for each of the treatment groups. Scale bars, 200 μm (**b**) and 1 cm (**i**). Data are mean  $\pm$  s.e.m. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$  (**c**, **f** and **g**; analysis of variance (ANOVA); Dunnett test (compared to wild type)).  $P$  values in **j** are given numerically, log-rank (Mantel–Cox) test;  $n \geq 8$  for each group.

(Fig. 4b, c). We next assessed whether EDNRB<sup>-/-</sup> ENCs display major differences in cell proliferation or survival as compared to wild-type ENCs. At 24 h after replating (time point used for scratch assay), we did not observe differences in either assay (Extended Data Fig. 7d, e). By 72 h, EDNRB<sup>-/-</sup> ENCs showed reduced proliferation while cell viability remained unaffected (Extended Data Fig. 7d, e).

We next carried out a small molecule screen to identify compounds capable of rescuing the migration defects observed in EDNRB<sup>-/-</sup> ENC precursors (Fig. 4d and Extended Data Fig. 8a). We screened 1,280 compounds (Prestwick Chemical Library) consisting of FDA-approved drugs. Results were binned into compounds with no effect, toxic effects, modest effects or strong effects (Fig. 4d and Extended Data Fig. 8a–c). The hits were further validated by repeating the scratch assay under non-high-throughput-screening conditions (Extended Data Fig. 8d). Compounds with strong effects were chosen for further follow-up. Among those validated compounds, we focused on the molecule pepstatin A, which showed a dose-dependent rescue of the migration defect

in EDNRB<sup>-/-</sup> ES-cell-derived NC precursors (Fig. 4e). Pepstatin A is a known inhibitor of acid proteases<sup>22</sup>. Among potential pepstatin targets, we explored BACE2 because RNA-seq data showed upregulation in ES-cell-derived NC lineages (Extended Data Fig. 9a), and BACE2 had been recently shown to modulate the migration of NC derivatives in the developing zebrafish embryo<sup>23</sup>. To address whether BACE2 mediates the pepstatin A effect, we tested structurally unrelated small molecules targeting BACE2. Exposure to the BACE inhibitor IV rescued the migration defect in the scratch assay similar to pepstatin A (Fig. 4f). Furthermore, BACE2 knockdown confirmed rescue of the migration defect in EDNRB<sup>-/-</sup> cells (Fig. 4g and Extended Data Fig. 9b). Finally, we tested whether pepstatin A exposure *in vitro* is sufficient to rescue the *in vivo* migration behaviour of EDNRB<sup>-/-</sup> ENC precursors (Fig. 4h). ENC precursors derived from EDNRB<sup>-/-</sup> cells exhibit a significant *in vivo* migration defect after transplantation into the adult colon (Fig. 4i, j). EDNRB-null precursors pre-treated with pepstatin A for 72 h before transplantation showed a significant rescue of *in vivo* migration (Fig. 4i, j and Extended Data Fig. 9c). Interestingly, wild-type-derived ENC precursors treated with pharmacological inhibitors of EDNRB showed migration defects *in vitro* and *in vivo* (Extended Data Fig. 9d–f), further supporting a role for EDNRB in human ENC migration and HSCR.

Our study describes an efficient strategy to derive and prospectively purify ENC precursors from human ES cells. In agreement with studies in model organisms<sup>1,2</sup>, we demonstrate that human ES-cell-derived ENC gives rise to a broad range of neurotransmitter phenotypes characteristic of the ENS. The ability to model human ENS development *in vitro* should enable the large-scale production of specific human enteric neurons on demand. For example human PS-cell-derived enteric 5-HT neurons could serve as a tool to model gastrointestinal side effects of CNS-acting drugs such as Prozac<sup>1</sup>.

We focused primarily on potential cell therapeutic applications of ES-cell-derived ENC lineages in HSCR. One of the most remarkable findings was the extensive *in vivo* migratory potential of human ENC precursors in the adult host colon. Future studies will have to define whether this ability is confined to early CD49D<sup>+</sup> cells or maintained at later neurogenic stages. Similarly, it will be important to look at long-term efficacy and safety. While most of our *in vivo* studies in NSG mice (a total of 102 animals grafted) were limited to a 5–6-week survival period, animals analysed at 3–4 months after transplantation showed comparable *in vivo* properties without evidence of tumour formation or graft-related adverse effects. The therapeutic potential of the cells is illustrated by their ability to rescue *Ednrb*<sup>s-l/s-l</sup> mice. Future studies will be required to address mechanisms of the graft-mediated host rescue. The potential for widespread engraftment may eventually enable permanent, bona fide repair of the aganglionic portions of the gut. However, given the rapid action of the cells in preventing death of HSCR mice, it appears unlikely that rescue was mediated by functional integration of the cells. Future studies should also address alternative mechanisms such as cytokine release, immunomodulation or changes in barrier function for contributing to the therapeutic effect.

The identification of pepstatin A and BACE2 inhibition in rescuing HSCR-related migration defects represents proof-of-concept for the use of human ES-cell-derived ENC precursors in drug discovery. The mechanism of BACE2 inhibition on migration remains to be determined. Possible targets of BACE proteases include neuregulins and ErbB receptors previously implicated in NC migration<sup>24,25</sup>. One obvious future direction is testing the therapeutic potential of BACE inhibitors in mouse models of HSCR. Such a strategy could enable the prevention of aganglionosis during pregnancy or enable repair of postnatal enteric neuron function. In the current study, we focused on combining pepstatin A as a neoadjuvant treatment to promote migration of EDNRB<sup>-/-</sup> ENC precursors. An important next question is whether cells pre-treated with pepstatin A are capable of rescuing lethality or other disease-associated phenotypes in HSCR mice. In conclusion, our work presents a powerful strategy to access human ENS lineages for exploring the second brain<sup>1</sup> in human health and disease and for developing novel cell- and drug-based therapies for HSCR.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 11 February; accepted 23 December 2015.**

**Published online 10 February 2016.**

- Gershon, M. *The Second Brain - A Groundbreaking New Understanding of Nervous Disorders of the Stomach and Intestine* (Harper Collins, 1999).
- Heanue, T. A. & Pachnis, V. Enteric nervous system development and Hirschsprung's disease: advances in genetic and stem cell studies. *Nature Rev. Neurosci.* **8**, 466–479 (2007).
- Mica, Y., Lee, G., Chambers, S. M., Tomishima, M. J. & Studer, L. Modeling neural crest induction, melanocyte specification, and disease-related pigmentation defects in hESCs and patient-specific iPSCs. *Cell Rep.* **3**, 1140–1152 (2013).
- Chan, K. K. *et al.* Hoxb3 vagal neural crest-specific enhancer element for controlling enteric nervous system development. *Dev. Dyn.* **233**, 473–483 (2005).
- Fu, M., Lui, V. C., Sham, M. H., Cheung, A. N. & Tam, P. K. HOXB5 expression is spatially and temporally regulated in human embryonic gut during neural crest cell colonization and differentiation of enteric neuroblasts. *Dev. Dyn.* **228**, 1–10 (2003).
- Wichterle, H., Lieberam, I., Porter, J. A. & Jessell, T. M. Directed differentiation of embryonic stem cells into motor neurons. *Cell* **110**, 385–397 (2002).
- Lee, G. *et al.* Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **461**, 402–406 (2009).
- Chalazonitis, A. *et al.* Neurotrophin-3 induces neural crest-derived cells from fetal rat gut to develop *in vitro* as neurons or glia. *J. Neurosci.* **14**, 6571–6584 (1994).
- Laflamme, M. A. *et al.* Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nature Biotechnol.* **25**, 1015–1024 (2007).
- Steinbeck, J. A. *et al.* Functional connectivity under optogenetic control allows modeling of human neuromuscular disease. *Cell Stem Cell* **18**, 134–143 (2016).
- Barthel, E. R. *et al.* Tissue engineering of the intestine in a murine model. *J. Vis. Exp.* **70**, e4279 (2012).
- Di Lorenzo, C., Solzi, G. F., Flores, A. F., Schwankovsky, L. & Hyman, P. E. Colonic motility after surgery for Hirschsprung's disease. *Am. J. Gastroenterol.* **95**, 1759–1764 (2000).
- Hotta, R., Natarajan, D., Burns, A. J. & Thapar, N. Stem cells for GI motility disorders. *Curr. Opin. Pharmacol.* **11**, 617–623 (2011).
- Schäfer, K. H., Micci, M. A. & Pasricha, P. J. Neural stem cell transplantation in the enteric nervous system: roadmaps and roadblocks. *Neurogastroenterol. Motil.* **21**, 103–112 (2009).
- Hotta, R. *et al.* Transplanted progenitors generate functional enteric neurons in the postnatal colon. *J. Clin. Invest.* **123**, 1182–1191 (2013).
- Garipey, C. E., Cass, D. T. & Yanagisawa, M. Null mutation of endothelin receptor type B gene in spotting lethal rats causes aganglionic megacolon and white coat color. *Proc. Natl Acad. Sci. USA* **93**, 867–872 (1996).
- Kruger, G. M. *et al.* Temporally distinct requirements for endothelin receptor B in the generation and migration of gut neural crest stem cells. *Neuron* **40**, 917–929 (2003).
- Tam, P. K. & Garcia-Barcelo, M. Genetic basis of Hirschsprung's disease. *Pediatr. Surg. Int.* **25**, 543–558 (2009).
- Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Chakravarti, A. Endothelin receptor-mediated signaling in hirschsprung disease. *Hum. Mol. Genet.* **5**, 303–307 (1996).
- Zhang, Y., Kim, T. H. & Niswander, L. Phactr4 regulates directional migration of enteric neural crest through PP1, integrin signaling, and cofilin activity. *Genes Dev.* **26**, 69–81 (2012).
- Yoshida, H. *et al.* Pepstatin A, an aspartic proteinase inhibitor, suppresses RANKL-induced osteoclast differentiation. *J. Biochem.* **139**, 583–590 (2006).
- Haas, H. A. Extending the search for folk personality constructs: the dimensionality of the personality-relevant proverb domain. *J. Pers. Soc. Psychol.* **82**, 594–609 (2002).
- Torii, T. *et al.* *In vivo* knockdown of ErbB3 in mice inhibits Schwann cell precursor migration. *Biochem. Biophys. Res. Commun.* **452**, 782–788 (2014).
- Wakatsuki, S., Araki, T. & Sehara-Fujisawa, A. Neuregulin-1/glia growth factor stimulates Schwann cell migration by inducing  $\alpha 5 \beta 1$  integrin-ErbB2-focal adhesion kinase complex formation. *Genes Cells* **19**, 66–77 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank K. Manova, M. Tomishima and A. Viale for technical assistance and G. Lee for human ES-cell-based PHOX2B-GFP reporter line. We are grateful for technical support provided by H. S. Ralph, K. Brodie and the MSKCC Flow Cytometry Core. We would also like to thank J.-F. Brunet for providing us with the PHOX2A antibody and P. Frykman for sharing their *EDNRB<sup>tm1Ywa</sup>/Fryk* HSCR mouse strain. The work was supported by the Starr Foundation and by NYSTEM contract C026446 to L.S. and by P30 CA008748; by grant NS15547 from the NINDS to M.D.G. and by grants RN200946-1 and RN3-06425 from the California Institute for Regenerative Medicine (CIRM) to T.C.G.; by TRI-SCI 2014-030 to L.S. and S.C., and by the New York Stem Cell Foundation (R-103) and NIDDK (DP2 DK098093-01) to S.C. S.C. is a New York Stem Cell Foundation – Robertson Investigator. J.A.S. was supported by a DFG fellowship. We would like to thank M. Tomishima and V. Tabar for comments on the manuscript.

**Author Contributions** F.F.: design and conception of the study, writing of manuscript, maintenance, directed differentiation and gene targeting of human PS cells, establishing of co-culture assays, cellular/molecular assays, histological analyses, small molecule screen. J.A.S.: establishing optogenetic human ES cell line and *in vitro* connectivity assays. S.K.: design and execution of chick transplantation assays. J.T.: RNA-seq experimental design and data analysis. B.Z.: design and data quantification for use of scratch assay. S.K.: biochemical analysis of *Ednrb*<sup>-/-</sup> cells. N.Z.: design of scratch assay. Y.M.: chick transplantation of MNCs. W.E.-N.: tissue-engineered intestine assay. E.D.S.: design and execution of colon injection assays in NSG and EDNRB mutant mice. H.Z.: colon injection assays in NSG and EDNRB mutant mice. M.G.: study design and data interpretation. T.G.: design and execution of tissue-engineered intestine assays. S.C.: design and interpretation of small molecule screen and follow-up experiment. L.S.: design and conception of the study, data interpretation, writing of manuscript.

**Author Information** RNA-seq data have been deposited in the Gene Expression Omnibus (GEO) under accession GSE66148. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.S. ([studerl@mskcc.org](mailto:studerl@mskcc.org)).

## METHODS

**Culture of undifferentiated human ES cells.** Human ES cell line H9 (WA-09) and derivatives (*SOX10::GFP*; *SYN::ChR2-EYFP*; *SYN::EYFP*; *PHOX2B::GFP*; *EF1::RFP* *Ednrb*<sup>−/−</sup>) as well as two independent human iPSC cell lines (healthy and familial dysautonomia, Sendai-based, OMSK (Cytotune)) were maintained on mouse embryonic fibroblasts (Global Stem) in knockout serum replacement (KSR; Life Technologies, 10828-028) containing human ES cell medium as described previously<sup>7</sup>. Cells were subjected to mycoplasma testing at monthly intervals and short tandem repeats (STR) profiled to confirm cell identity at the initiation of the study.

**Neural crest induction.** Human ES cells were plated on matrigel (BD Biosciences, 354234)-coated dishes (10<sup>5</sup> cells cm<sup>−2</sup>) in ES cell medium containing 10 nM FGF2 (R&D Systems, 233-FB-001MG/CF). Differentiation was initiated in KSR medium (knockout DMEM plus 15% KSR (Life Technologies, 10828-028), L-glutamine (Life Technologies, 25030-081), NEAA (Life Technologies, 11140-050)) containing LDN193189 (100 nM, Stemgent) and SB431542 (10 μM, Tocris). The KSR medium was gradually replaced with increasing amounts of N2 medium from day 4 to day 10 as described previously<sup>7</sup>. For CNC induction, cells were treated with 3 μM CHIR99021 (Tocris Bioscience, 4423) in addition to LDN193189 and SB431542 from day 2 to day 11. ENC differentiation involves additional treatment with retinoic acid (1 μM) from day 6 to day 11. For deriving MNCs, LDN193189 is replaced with BMP4 (10 nM, R&D, 314-BP) and EDN3 (10 nM, American Peptide company, 88-5-10B) from day 6 to day 11 (ref. 3). The differentiated cells are sorted for CD49D at day 11. CNS precursor control cells were generated by treatment with LDN193189 and SB431542 from day 0 to day 11 as previously described<sup>7</sup>. Throughout the manuscript, day 0 is the day the medium is switched from human ES cell medium to LDN193189 and SB431542 containing medium. Days of differentiation in text and figures refer to the number of days since the pluripotent stage (day 0).

**FACS and immunofluorescence analysis.** For immunofluorescence, the cells were fixed with 4% paraformaldehyde (Affymetrix-USB, 19943) for 20 min, then blocked and permeabilized using 1% bovine serum albumin (BSA) (Thermo Scientific, 23209) and 0.3% Triton X-100 (Sigma, T8787). The cells were then incubated in primary antibody solutions overnight at 4 °C and stained with fluorophore-conjugated secondary antibodies at room temperature for 1 h. The stained cells were then incubated with DAPI (1 ng ml<sup>−1</sup>, Sigma, D9542-5MG) and washed several times before imaging. For flow cytometry analysis, the cells are dissociated with Accutase (Innovative Cell Technologies, AT104) and fixed and permeabilized using BD Cytotfix/Cytoperm (BD Bioscience, 554722) solution, then washed, blocked and permeabilized using BD Perm/Wash buffer (BD Bioscience, 554723) according to manufacturer's instructions. The cells are then stained with primary (overnight at 4 °C) and secondary (30 min at room temperature) antibodies and analysed using a flow Cytometer (Flowjo software). A list of primary antibodies and working dilutions is provided in Supplementary Table 4. The PHOX2A antibody was provided by J.-F. Brunet (rabbit, 1:800 dilution).

**In ovo transplants.** Fertilized eggs (from Charles River Farms) were incubated at 37 °C for 50 h before injections. A total of 2 × 10<sup>5</sup> CD49D-sorted, RFP-labelled NC cells were injected into the intersomitic space of the vagal region of the embryos targeting a region between somite 2 and 6 (HH 14 embryo, 20–25 somite stage). The embryos were collected 36 h later for whole-mount epifluorescence and histological analyses.

**Gene expression analysis.** For RNA sequencing, total RNA was extracted using RNeasy RNA purification kit (Qiagen, 74106). For qRT-PCR assay, total RNA samples were reverse transcribed to cDNA using Superscript II Reverse Transcriptase (Life Technologies, 18064-014). qRT-PCR reactions were set up using QuantiTect SYBR Green PCR mix (Qiagen, 204148). Each data point represents three independent biological replicates.

**In vitro differentiation of ENC to enteric neurons.** ENC cells from the 11-day induction protocol were aggregated into 3D spheroids (5 million cells per well) in Ultra Low Attachment 6-well culture plates (Fisher Scientific, 3471) and cultured in Neurobasal (NB) medium supplemented with L-glutamine (Gibco, 25030-164), N2 (Stem Cell Technologies, 07156) and B27 (Life Technologies, 17504044) containing CHIR99021 (3 μM, Tocris Bioscience, 4423) and FGF2 (10 nM, R&D Systems, 233-FB-001MG/CF). After 4 days of suspension culture, the spheroids are plated on poly-ornithine/laminin/fibronectin (PO/LM/FN)-coated dishes (prepared as described previously<sup>26</sup>) in neurobasal (NB) medium supplemented with L-glutamine (Gibco, 25030-164), N2 (Stem Cell Technologies, 07156) and B27 (Life Technologies, 17504044) containing GDNF (25 ng ml<sup>−1</sup>, Peprotech, 450-10) and ascorbic acid (100 μM, Sigma, A8960-5G). The ENC precursors migrate out of the plated spheroids and differentiate into neurons in 1–2 weeks. The cells were fixed for immunostaining or collected for gene expression analysis at days 25, 40 and 60 of differentiation.

**Induction of SMCs.** Mesoderm specification is carried out in STEMPro-34 (Gibco, 10639-011) medium. The ES cells are subjected to activin A treatment (100 ng ml<sup>−1</sup>, R&D, 338-AC-010) for 24 h followed by BMP4 treatment (10 ng ml<sup>−1</sup>, R&D, 314-BP) for 4 days<sup>9</sup>. The cells are then differentiated into SMC progenitors by treatment with PDGF-BB (5 ng ml<sup>−1</sup>, Peprotech, 100-14B), TGFβ3 (5 ng ml<sup>−1</sup>, R&D systems, 243-B3-200) and 10% FBS. The SMC progenitors are expandable in DMEM supplemented with 10% FBS.

**ENC-SMC co-culture.** The SMC progenitors were plated on PO/LM/FN-coated culture dishes (prepared as described previously<sup>26</sup>) 3 days before addition of ENC-derived neurons. The neurons were dissociated (using accutase, Innovative Cell Technologies, AT104) at day 30 of differentiation and plated onto the SMC monolayer cultures. The culture is maintained in neurobasal (NB) medium supplemented with L-glutamine (Gibco, 25030-164), N2 (Stem Cell Technologies, 07156) and B27 (Life Technologies, 17504044) containing GDNF (25 ng ml<sup>−1</sup>, Peprotech, 450-10) and ascorbic acid (100 μM, Sigma, A8960-5G). Functional connectivity was assessed at 8–16 weeks of co-culture.

**Pharmacological and optogenetic stimulations of co-cultured SMCs.** SMC-only and SMC-ENC-derived neuron co-cultures were subjected to acetylcholine chloride (50 μM, Sigma, A6625), carbamoylcholine chloride (10 μM, Sigma, C4382) and KCl (55 mM, Fisher Scientific, BP366-500) treatment, 3 months after initiating the co-culture. Optogenetic stimulations were performed using a 450-nm pigtailed diode pumped solid state laser (OEM Laser, PSU-III LED, OEM Laser Systems, Inc.) achieving an illumination between 2 and 4 mW mm<sup>−2</sup>. The pulse width was 4 ms and stimulation frequencies ranged from 2 to 10 Hz. For the quantification of movement, images were assembled into a stack using Metamorph software and regions with high contrast were identified (labelled yellow in Supplementary Fig. 5). The movement of five representative high-contrast regions per field was automatically traced (Metamorph software). Data are presented in kintograms as movement in pixels in x and y direction (distance) with respect to the previous frame.

**Generation of chimaeric tissue-engineered colon.** We used the previously described method for generation of tissue-engineered colon<sup>11</sup>. In brief, the donor colon tissue was collected and digested into organoid units using dispase (Life Technologies, 17105-041) and collagenase type 1 (Worthington, CLS-1). The organoid units were then mixed immediately (without any *in vitro* culture) with CD49D-purified human ES-cell-derived ENC precursors (day 15 of differentiation) and seeded onto biodegradable polyglycolic acid scaffolds (2-mm sheet thickness, 60 mg cm<sup>−3</sup> bulk density; porosity >95%, Concordia Fibres) shaped into 2 mm long tubes with poly-L lactide (PLLA) (Durect Corporation). The seeded scaffolds were then placed onto and wrapped in the greater omentum of the adult (>2 months old) NSG mice. Just before the implantation, these mice were irradiated with 350 cGy. The seeded scaffolds were differentiated into colon-like structures inside the omentum for 4 additional weeks before they were surgically removed for tissue analysis.

**Transplantation of ENC precursors in adult colon.** All mouse procedures were performed following NIH guidelines, and were approved by the local Institutional Animal Care and Use Committee (IACUC), the Institutional Biosafety Committee (IBC) as well as the Embryonic Stem Cell Research Committee (ESRCO). We used 3–6-week-old male NSG (NOD.Cg-Prkdc<sup>scid</sup> Il2rg<sup>tm1Wjl/SzJ</sup>) mice or 2–3-week-old *Ednrb*<sup>−/s-1</sup> (SSL/Le) mice<sup>27</sup> (*n* = 12, 6 male, 6 female) for these studies. Animal numbers were based on availability of homozygous hosts and on sufficient statistical power to detect large effects between treatment versus control (*Ednrb*<sup>−/s-1</sup>) as well as for demonstrating robustness of migration behaviour (NSG). Animals were randomly selected for the various treatment models (NSG and *Ednrb*<sup>−/s-1</sup>) but assuring for equal distribution of male/female ratio in each group (*Ednrb*<sup>−/s-1</sup>). All *in vivo* experiments were performed in a blinded manner. Animals were anaesthetized with isoflurane (1%) throughout the procedure, a small abdominal incision was made, abdominal wall musculature lifted and the caecum is exposed and exteriorized. Warm saline is used to keep the caecum moist. Then 20 μl of cell suspension (2–4 million RFP<sup>+</sup> CD49D-purified human ES-cell-derived ENC precursors) in 70% Matrigel (BD Biosciences, 354234) in PBS or 20 μl of 70% Matrigel in PBS only (control-grafted animals) were slowly injected into the caecum (targeting the muscle layer) using a 27-gauge needle. Use of 70% matrigel as carrier for cell injection assured that the cells stayed in place after the injection and prevented backflow into the peritoneum. After injection that needle was withdrawn, and a Q-tip was placed over the injection site for 30 s to prevent bleeding. The caecum was returned to the abdominal cavity and the abdominal wall was closed using 4-0 vicryl and a taper needle in an interrupted suture pattern and the skin was closed using sterile wound clips. After wound closure animals were put on paper on top of their bedding and attended until conscious and preferably eating and drinking. The tissue was collected at different time points (ranging from two weeks to four months) after transplantation for histological analysis. *Ednrb*<sup>−/s-1</sup> mice were immunosuppressed by daily injections of cyclosporine (10 mg kg<sup>−1</sup> i.p., Sigma, 30024).



**Whole-mount fluorescence imaging and histology.** The collected colon samples were fixed in 4% paraformaldehyde at 4 °C overnight before imaging. Imaging is performed using Maestro fluorescence imaging system (Cambridge Research and Instrumentation). The tissue samples were incubated in 30% sucrose (Fisher Scientific, BP220-1) solutions at 4 °C for 2 days, and then embedded in OCT (Fisher Scientific, NC9638938) and cryosectioned. The sections were then blocked with 1% BSA (Thermo Scientific, 23209) and permeabilized with 0.3% Triton X-100 (Sigma, T8787). The sections are then stained with primary antibody solution at 4 °C overnight and fluorophore-conjugated secondary antibody solutions at room temperature for 30 min. The stained sectioned were then incubated with DAPI (1 ng ml<sup>-1</sup>, Sigma, D9542-5MG) and washed several times before they were mounted with Vectashield Mounting Medium (vector, H1200) and imaged using fluorescent (Olympus IX70) or confocal microscopes (Zeiss SP5).

**Total gastrointestinal transit time.** Mice are gavaged with 0.3 ml of dye solution containing 6% carmine (Sigma, C1022-5G), 0.5% methylcellulose (Sigma, 274429-5G) and 0.9 NaCl, using a #24 round-tip feeding needle. The needle was held inside the mouse oesophagus for a few seconds after gavage to prevent regurgitation. After 1 h, the stool colour was monitored for gavaged mice every 10 min. For each mouse, total gastrointestinal transit time is between the time of gavage and the time when red stool is observed.

**Gene targeting.** The double nickase CRISPR/Cas9 system<sup>28</sup> was used to target the EDNRB locus in EF1-RFP H9 human ES cells. Two guide RNAs were designed (using the CRISPR design tool; <http://crispr.mit.edu/>) to target the coding sequence with PAM targets ~20 base pairs apart (qRNA #1 target specific sequence: 5'-AAGTCTGTGCGGACGCGCCCTGG-3', RNA #2 target specific sequence: 5'-CCAGATCCGCGACAGGCCGAGG-3'). The cells were transfected with guide RNA constructs and GFP-fused Cas9-D10A nickase. The GFP-expressing cells were FACS purified 24 h later and plated in low density (150 cells cm<sup>-2</sup>) on mouse embryonic fibroblasts. The colonies were picked 7 days later and passaged twice before genomic DNA isolation and screening. The targeted region of EDNRB gene was PCR amplified (forward primer: 5'-ACGCCTTCTGGAGCAGGTAG-3', reverse primer: 5'-GTCAGGCGGGAAGCCTCTCT-3') and cloned into Zero Blunt TOPO vector (Invitrogen, 450245). To ensure that both alleles (from each ES cell colony) are represented and sequenced, we picked 10 bacterial clones (for each ES cell clone) for plasmid purification and subsequent sequencing. The clones with bi-allelic nonsense mutations were expanded and differentiated for follow-up assays.

**Migration assay.** The ENC cells are plated on PO/LM/FN coated (prepared as described previously<sup>26</sup>) 96-well or 48-well culture plates (30,000 cm<sup>-2</sup>). After 24 h, the culture lawn is scratched manually using a pipette tip. The cells are given an additional 24–48 h to migrate into the scratch area and fixed for imaging and quantification. The quantification is based on the percentage of the nuclei that are located in the scratch area after the migration period. The scratch area is defined using a reference well that was fixed immediately after scratching. Migration of cells was quantified using the open source data analysis software KNIME<sup>29</sup> (<http://knime.org>) with the 'quantification in ROI' plug-in as described in detail elsewhere<sup>30</sup>.

**Proliferation assay.** To quantify proliferation, FACS-purified ENC cells were assayed using CyQUANT NF cell proliferation Assay Kit (Life Technologies, C35006) according to manufacturer's instructions. In brief, to generate a standard, cells were plated at various densities and stained using the fluorescent DNA binding dye reagent. Total fluorescence intensity was then measured using a plate reader (excitation at 485 nm and emission detection at 530 nm). After determining the linear range, the CD49D<sup>+</sup> wild-type and EDNRB<sup>-/-</sup> ENC precursors were plated (6,000 cell cm<sup>-2</sup>) and assayed at 0, 24, 48 and 72 h. The cells were cultured in neurobasal (NB) medium supplemented with L-glutamine (Gibco, 25030-164), N2 (Stem Cell Technologies, 07156) and B27 (Life Technologies, 17504044) containing CHIR99021 (3 μM, Tocris Bioscience, 4423) and FGF2 (10 nM, R&D Systems, 233-FB-001MG/CF) during the assay.

**Viability assay.** To monitor the viability of wild-type and EDNRB<sup>-/-</sup> ENC precursors, cells were assayed for lactate dehydrogenase (LDH) activity using CytoTox 96 cytotoxicity assay kit (Promega, G1780). In brief, the cells are plated in 96-well plates at 30,000 cm<sup>-2</sup>. The supernatant and the cell lysate is collected 24 h later and assayed for LDH activity using a plate reader (490 nm absorbance). Viability is calculated by dividing the LDH signal of the lysate by total LDH signal (from lysate plus supernatant). The cells were cultured in neurobasal (NB) medium supplemented with L-glutamine (Gibco, 25030-164), N2 (Stem Cell Technologies, 07156)

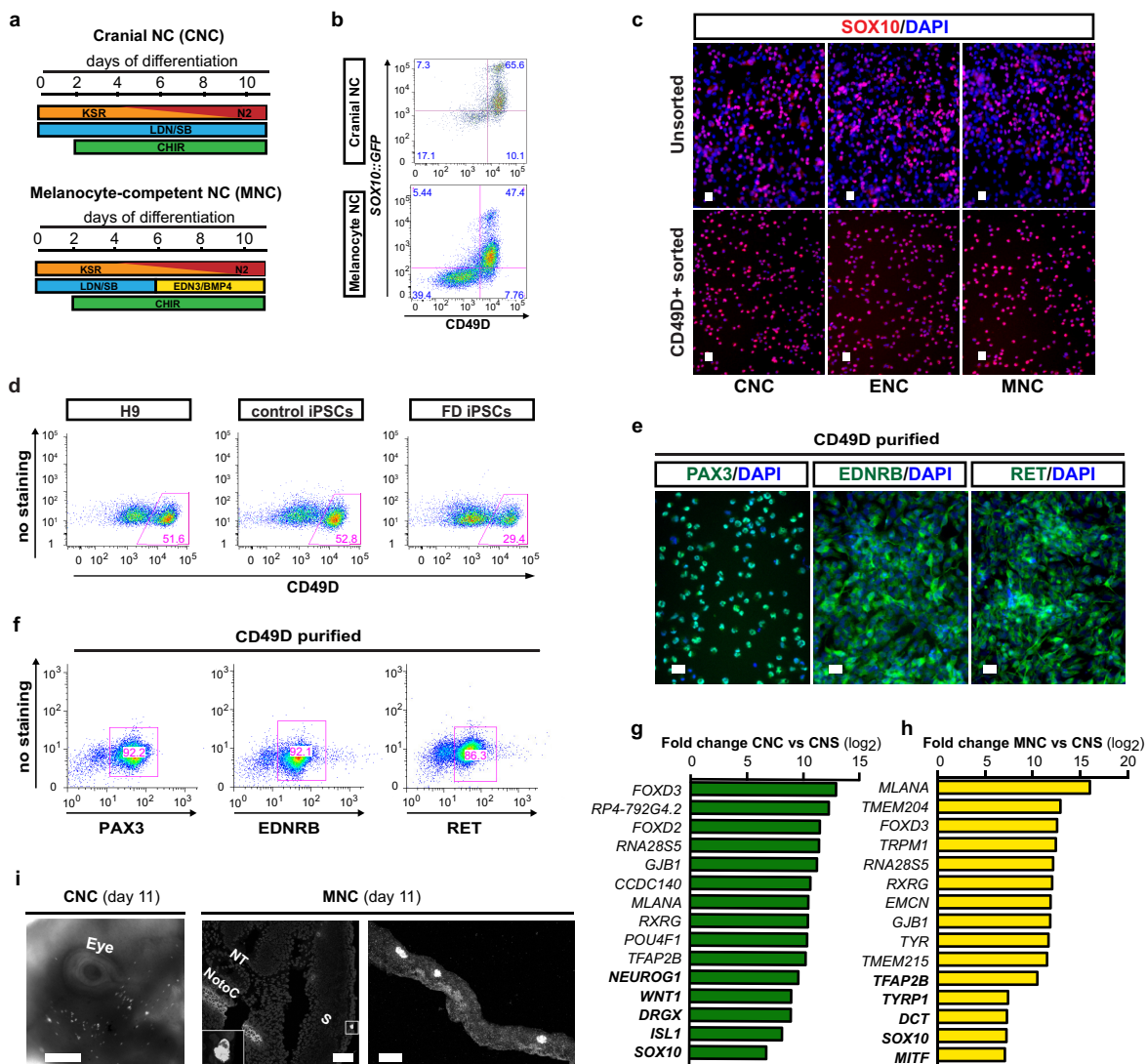
and B27 (Life Technologies, 17504044) containing CHIR99021 (3 μM, Tocris Bioscience, 4423) and FGF2 (10 nM, R&D Systems, 233-FB-001MG/CF) during the assay.

**High-throughput screening.** The chemical compound screening was performed using the Prestwick Chemical Library. The ENC cells were plated in 96-well plates (30,000 cm<sup>-2</sup>) and scratched manually 24 h before addition of the compounds. The cells were treated with two concentrations of the compounds (10 μM and 1 μM). The plates were fixed 24 h later for total plate imaging. The compounds were scored based on their ability to promote filling of the scratch in 24 h. The compounds that showed toxic effects (based on marked reduction in cell numbers assessed by DAPI staining) were scored 0, compounds with no effects were scored 1, compounds with moderate effects were scored 2, and compounds with strong effects (that resulted in complete filling of the scratch area) were scored 3 and identified as hit compounds. The hits were further validated to ensure reproducibility. The cells were treated with various concentrations of the selected hit compound (pepstatin A) for dose response analysis. The optimal dose (10 μM based on optimal response and viability) was used for follow-up experiments. For the pre-treatment experiments, cells were CD49D purified at day 11 and treated with pepstatin A from day 12 to day 15 followed by transplantation into the colon wall of NSG mice. The cells were cultured in neurobasal (NB) medium supplemented with L-glutamine (Gibco, 25030-164), N2 (Stem Cell Technologies, 07156) and B27 (Life Technologies, 17504044) containing CHIR99021 (3 μM, Tocris Bioscience, 4423) and FGF2 (10 nM, R&D Systems, 233-FB-001MG/CF) during the assay.

**BACE2 inhibition and knockdown.** To inhibit BACE2, the ENC precursors were treated with 1 μM β-secretase inhibitor IV (CAS 797035-11-1; Calbiochem). To knockdown BACE2, cells were dissociated using accutase (Innovative Cell Technologies, AT104) and reverse-transfected (using Lipofectamine RNAiMAX-Life Technologies, 13778-150) with an siRNA pool (SMARTpool: ON-TARGETplus BACE2 siRNA, Dharmacon, L-003802-00-0005) or four different individual siRNAs (Dharmacon, LQ-003802-00-0002, 2 nmol). The knockdown was confirmed by qRT-PCR measurement of BACE2 mRNA levels in cells transfected with the BACE2 siRNAs versus the control siRNA pool (ON-TARGETplus Non-targeting Pool, Dharmacon, D-001810-10-05). The transfected cells were scratched 24 h after plating and fixed 48 h later for migration quantification. The cells were cultured in neurobasal (NB) medium supplemented with L-glutamine (Gibco, 25030-164), N2 (Stem Cell Technologies, 07156) and B27 (Life Technologies, 17504044) containing CHIR99021 (3 μM, Tocris Bioscience, 4423) and FGF2 (10 nM, R&D Systems, 233-FB-001MG/CF) during the assay.

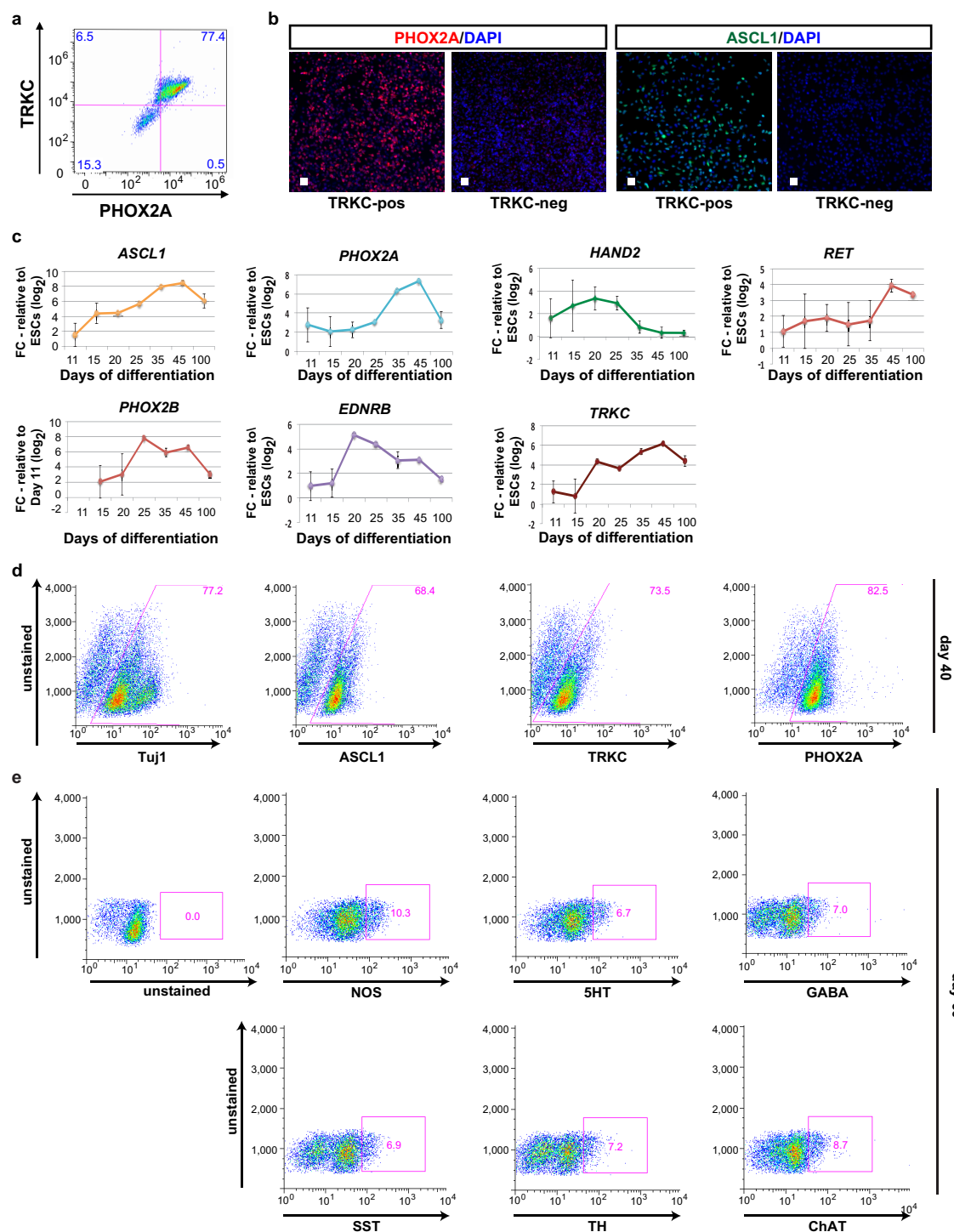
**Statistical analysis.** Data are presented as mean ± s.e.m. and were derived from at least three independent experiments. Data on replicates (*n*) is given in figure legends. Statistical analysis was performed using the Student's *t*-test (comparing two groups) or ANOVA with Dunnett test (comparing multiple groups against control). Distribution of the raw data approximated normal distribution (Kolmogorov–Smirnov normality test) for data with sufficient number of replicates to test for normality. Survival analysis was performed using a log-rank (Mantel–Cox) test. Z-scores for primary hits were calculated as  $Z = (x - \mu)/\sigma$ , in which *x* is the migration score value and is 3 for all hit compounds;  $\mu$  is the mean migration score value, and  $\sigma$  is the standard deviation for all compounds and DMSO controls (*n* = 224).

26. Zeltner, N., Lafaille, F. G., Fattahi, F. & Studer, L. Feeder-free derivation of neural crest progenitor cells from human pluripotent stem cells. *J. Vis. Exp.* **87**, e51609 (2014).
27. Hosoda, K. *et al.* Targeted and natural (piebald-lethal) mutations of endothelin-B receptor gene produce megacolon associated with spotted coat color in mice. *Cell* **79**, 1267–1276 (1994).
28. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
29. Berthold, M. *et al.* in *Data Analysis, Machine Learning and Applications Studies in Classification, Data Analysis, and Knowledge Organization* (eds Preisach, C., Burkhardt, H., Schmidt-Thieme, L. & Decker, R.) Ch. 38, 319–326 (Springer Berlin Heidelberg, 2008).
30. Dreser, N. *et al.* Grouping of histone deacetylase inhibitors and other toxicants disturbing neural crest migration by transcriptional profiling. *Neurotoxicology* **50**, 56–70 (2015).
31. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnol.* **27**, 275–280 (2009).



**Extended Data Figure 1 | Characterization of ES-cell-derived NC populations.** **a**, Schematic illustration of CNC and MNC induction protocols<sup>3</sup>. **b**, Flow cytometry for CD49D and SOX10–GFP in CNC and MNC cells. **c**, Immunofluorescence of unsorted and CD49D-sorted differentiated NC cells for SOX10. **d**, Flow cytometry for CD49D in CNCs derived from H9 human ES cells and control and familial dysautonomia (FD) human iPSC cells. **e**, **f**, Representative immunofluorescence images and flow cytometry in ES-cell-derived ENC for enteric precursor lineage marker at day 11. **g**, List of the top 10 and selected additional (bold) most

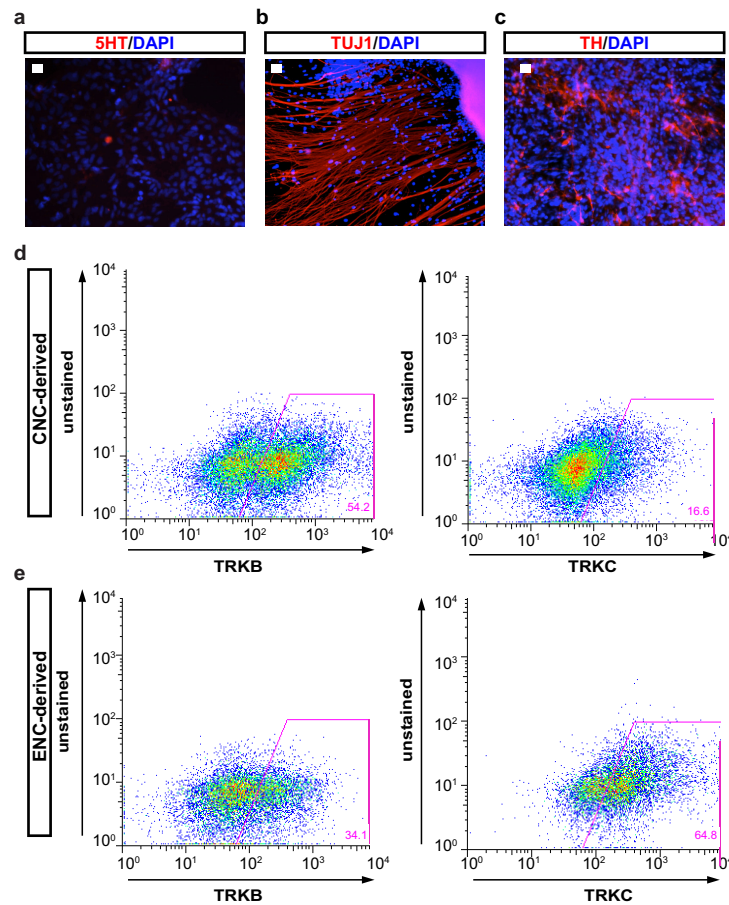
differentially expressed transcripts from the RNA-seq analysis of CNC compared to stage matched CNS precursors<sup>31</sup>. **h**, Lists of the top 10 and selected additional (bold) most differentially expressed transcripts from the RNA-seq analysis of MNC compared to stage-matched CNS precursors<sup>31</sup>. **i**, Distribution of CNC and MNC cells in developing chick embryos at 24–36 h after injection. Right, higher power image of the clusters of MNC cells in the developing surface ectoderm. NotoC, notochord; NT, neural tube; S, somite. Scale bars, 50 μm (**c**, **i**, middle), 25 μm (**e**, **i**, right) and 1 mm (**i**, left).



**Extended Data Figure 2 | Characterization of human ES-cell-derived enteric neural lineages.** **a**, Flow cytometry for TRKC and PHOX2A expression. **b**, Immunofluorescence for PHOX2A and ASCL1 for TRKC-positive and TRKC-negative ES-cell-derived ENC precursors. **c**, Time course qRT-PCR analysis of enteric lineage markers during more extended

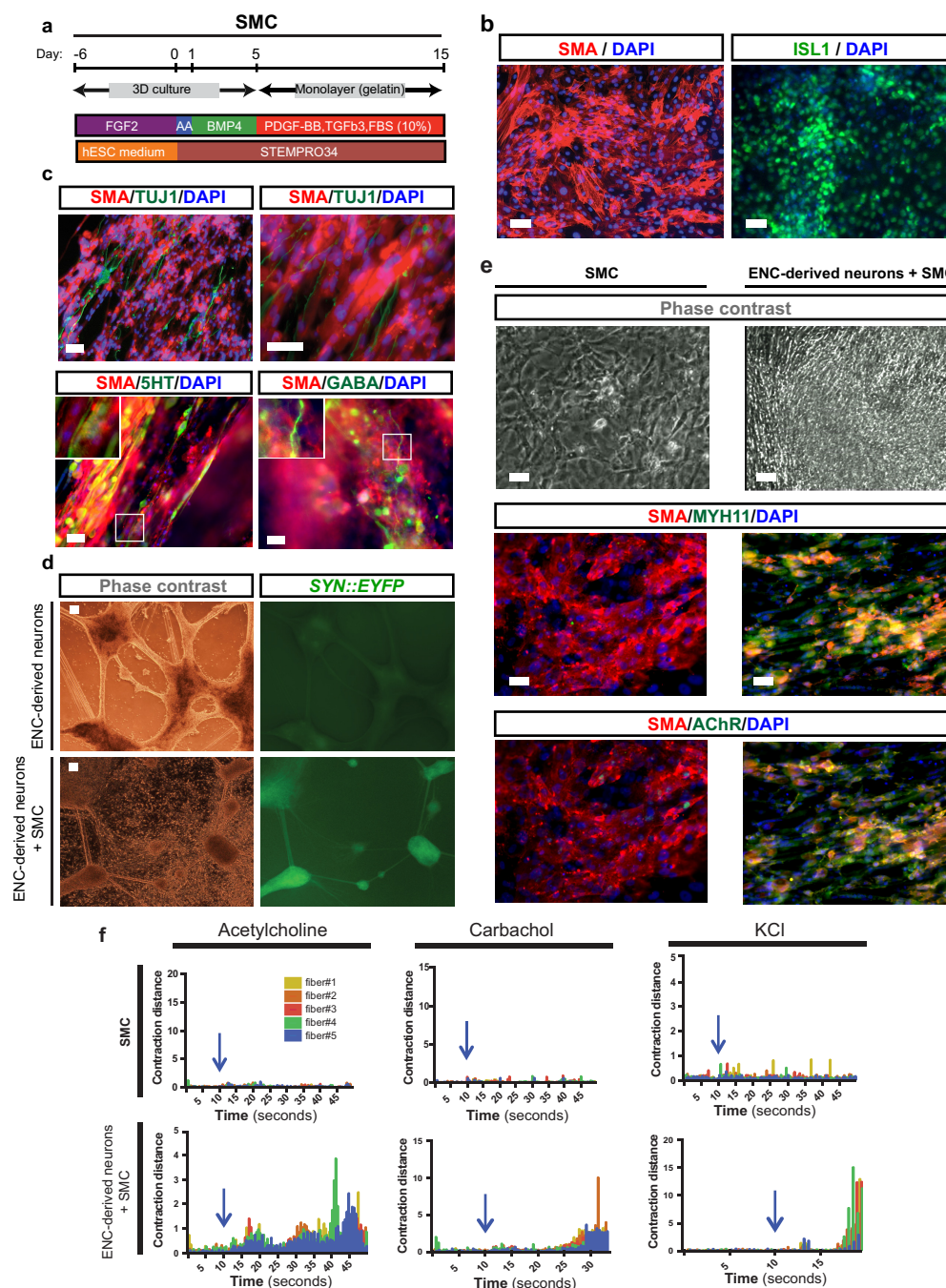
*in vitro* differentiation periods;  $n = 3$  independent experiments. **d**, **e**, Flow cytometric quantification of enteric neuron precursor and neuronal markers in ENC-derived neurons at days 40 and 60 of differentiation;  $n = 3$  independent experiments. Scale bars, 50  $\mu$ m.





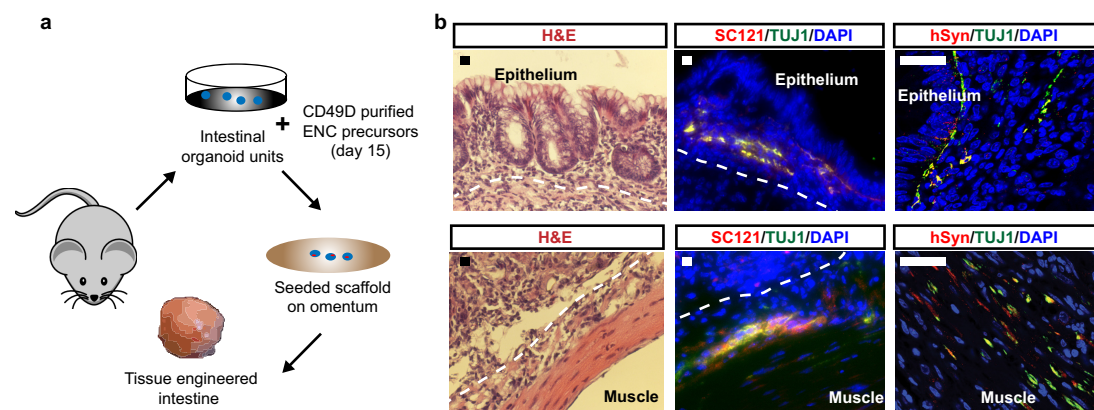
**Extended Data Figure 3 | CNC gives rise to neurons enriched in autonomic lineage.** **a–c,** Representative immunofluorescence images for expression of 5-HT, TUJ1 and tyrosine hydroxylase (TH) in CNC-derived neurons. In contrast to ENC-derived lineages, no serotonergic (5-HT<sup>+</sup>)

neurons were detected under cranial conditions despite the presence of many TUJ1<sup>+</sup> neurons and increased percentages of tyrosine-hydroxylase-positive cells. **d, e,** Flow cytometry for TRKB and TRKC under CNC and ENC conditions;  $n = 3$  independent experiments. Scale bars, 50  $\mu\text{m}$ .



**Extended Data Figure 4 | Functional characterization of ES-cell-derived enteric neurons in co-culture with SMCs.** **a**, Schematic illustration of SMC differentiation protocol. **b**, Immunofluorescence staining of SMC progenitors for SMA and ISL1. **c**, Association of various ENC-derived neuron subtypes with SMA<sup>+</sup> cells. **d**, SYN-eGFP expression in ENC-derived neurons at 40 days of co-culture with SMCs and stage-matched neurons in the absence of SMCs. **e**, Monoculture of SMCs versus

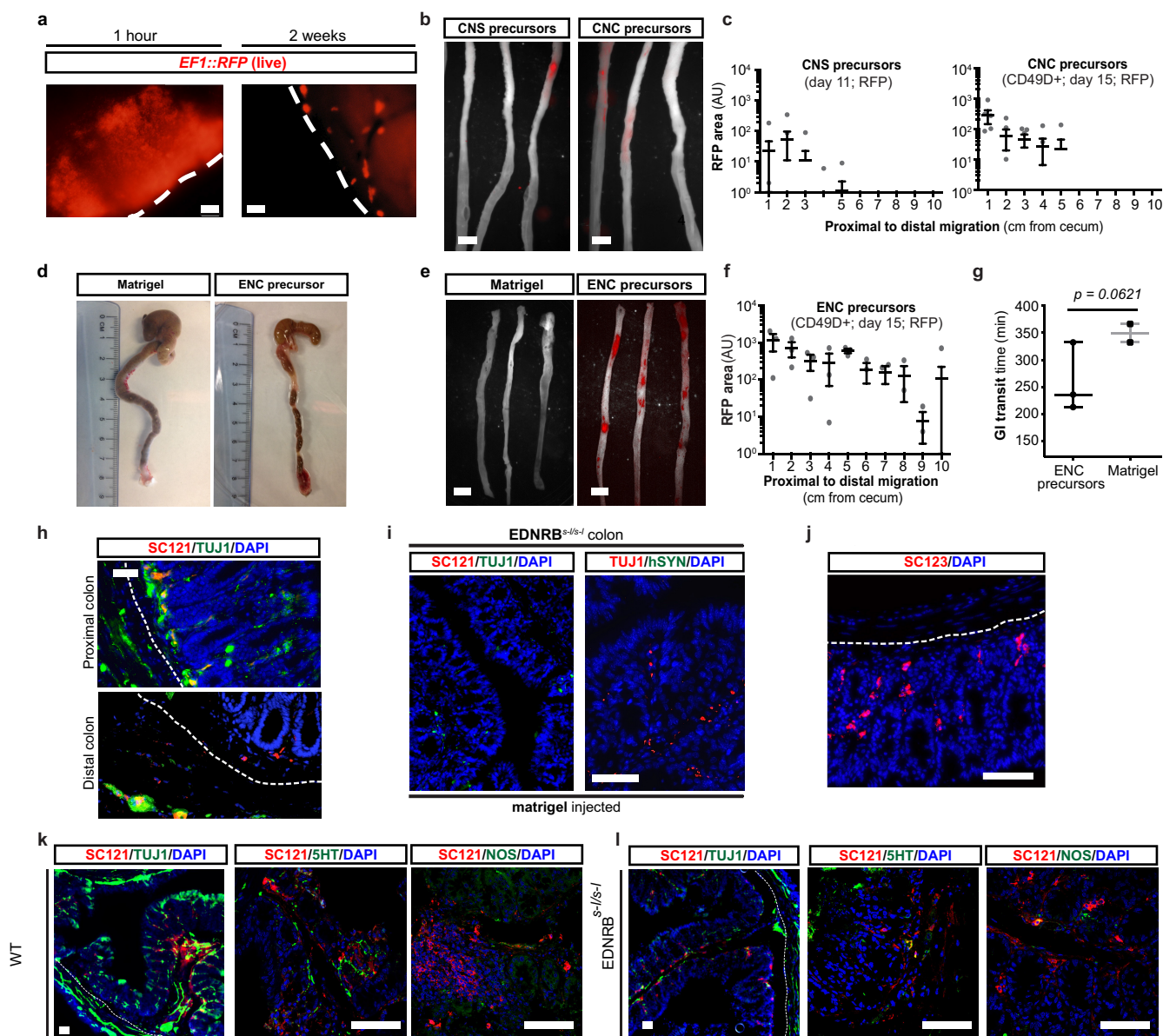
co-cultures of SMCs with ENC-derived neurons. Top, phase-contrast images showing morphological changes of SMCs in co-culture. Bottom, immunofluorescence staining of mature SMC markers MYH11 and AChR in monoculture of SMCs versus co-cultures of SMCs with ENC-derived neurons. **f**, Diagrams representing extent of contraction of SMC cultures. Arrows indicate the time of pharmacological stimulation. Scale bars, 50 μm (**b**, **c**, **e**) and 100 μm (**d**).



**Extended Data Figure 5 | Generation of tissue-engineered colon using human ES-cell-derived ENC precursors.** **a**, Schematic illustration for generation of tissue-engineered colon. **b**, Tissue engineered colon stained for human cytoplasmic marker SC121, TUJ1 and human-specific

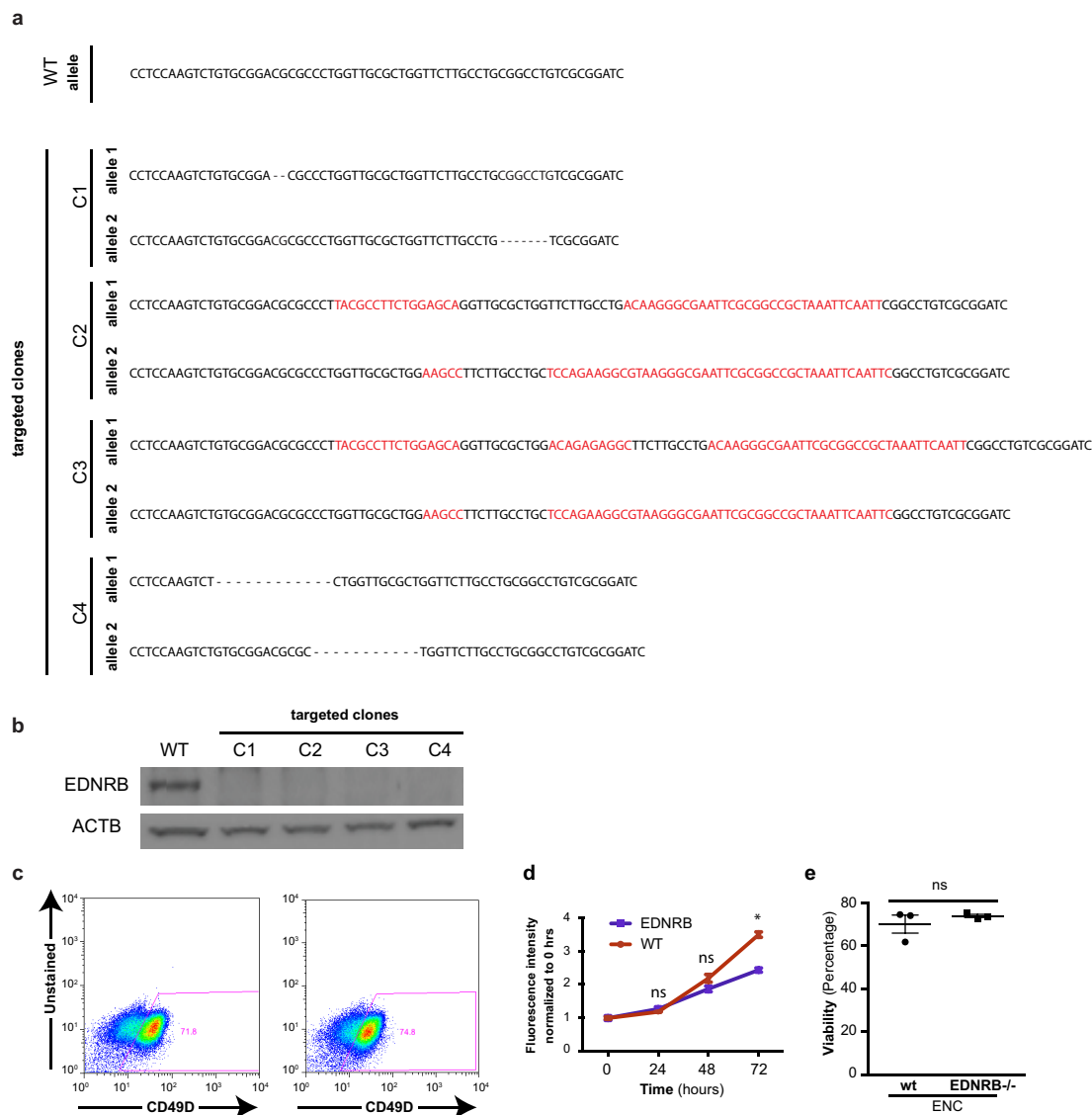
synaptophysin (hSyn). Dotted line shows approximate location of border between muscle and epithelial/submucosal-like layers. H&E, haematoxylin and eosin. Scale bars, 20  $\mu\text{m}$  (**b**, left and middle) and 40  $\mu\text{m}$  (**b**, right).





**Extended Data Figure 6 | Characterization of transplanted human ES-cell-derived ENC precursors in adult colon of NSG and *Ednrb*<sup>s-l/s-l</sup> mice.** **a**, Whole-mount microscopy of colon transplanted with RFP<sup>+</sup> CD49D-purified ES-cell-derived ENC precursors to track RFP expression at injection site, at 1 h after transplantation to ensure that cells were injected at proper location (left), and at 2 weeks to show dispersal of the cells and congregation of subset of cells into distinct clusters (right). The dashed lines indicate the outer border of the intact colon tissue. **b**, **c**, Whole-mount fluorescence imaging and quantification of migration of grafted RFP<sup>+</sup> ES-cell-derived CNS precursors and CD49D-purified CNC precursors inside the adult colon wall. **d**, Megacolon-like phenotype in control animals versus animals receiving ES-cell-derived ENC transplants. **e**, **f**, Whole-mount fluorescence imaging and quantification of migration of grafted RFP<sup>+</sup> CD49D purified ES-cell-derived ENC precursors in colon of *Ednrb*<sup>s-l/s-l</sup> mice. **g**, Total gastrointestinal transit times by carmine dye gavage of *Ednrb*<sup>s-l/s-l</sup> mice grafted with RFP<sup>+</sup> CD49D-purified ES-cell-derived ENC precursors versus Matrigel-only

grafted mice;  $n = 3$  for grafted animals,  $n = 2$  for Matrigel group. Note that  $n = 2$  for Matrigel group was due to the fact that nearly all Matrigel-injected animals died owing to their disease phenotype. **h**, Representative images of grafted ES-cell-derived ENC precursors at 3 months after transplantation into the colon of *Ednrb*<sup>s-l/s-l</sup> mice co-expressing TUJ1 and SC121. **i**, Immunofluorescence staining of cross sections of *Ednrb*<sup>s-l/s-l</sup> colons transplanted with Matrigel for SC121 and human-specific synaptophysin. **j**, Representative image of grafted ES-cell-derived ENC precursors at 3 months after transplantation into the colon of *Ednrb*<sup>s-l/s-l</sup> mice expressing human-specific GFAP (SC123). **k**, **l**, Representative images of grafted ES-cell-derived ENC precursors at 6 weeks after transplantation into the colon of NSG (wild type) and *Ednrb*<sup>s-l/s-l</sup> mice. The dashed lines indicate the border between submucosal and muscle layers. Scale bars, 200  $\mu$ m (**a**), 1 cm (**b**, **e**) and 100  $\mu$ m (**h**–**l**). AU, arbitrary units.  $P$  value for **g** is given numerically,  $t$ -test with Welch's correction;  $n = 3$  independent experiments.



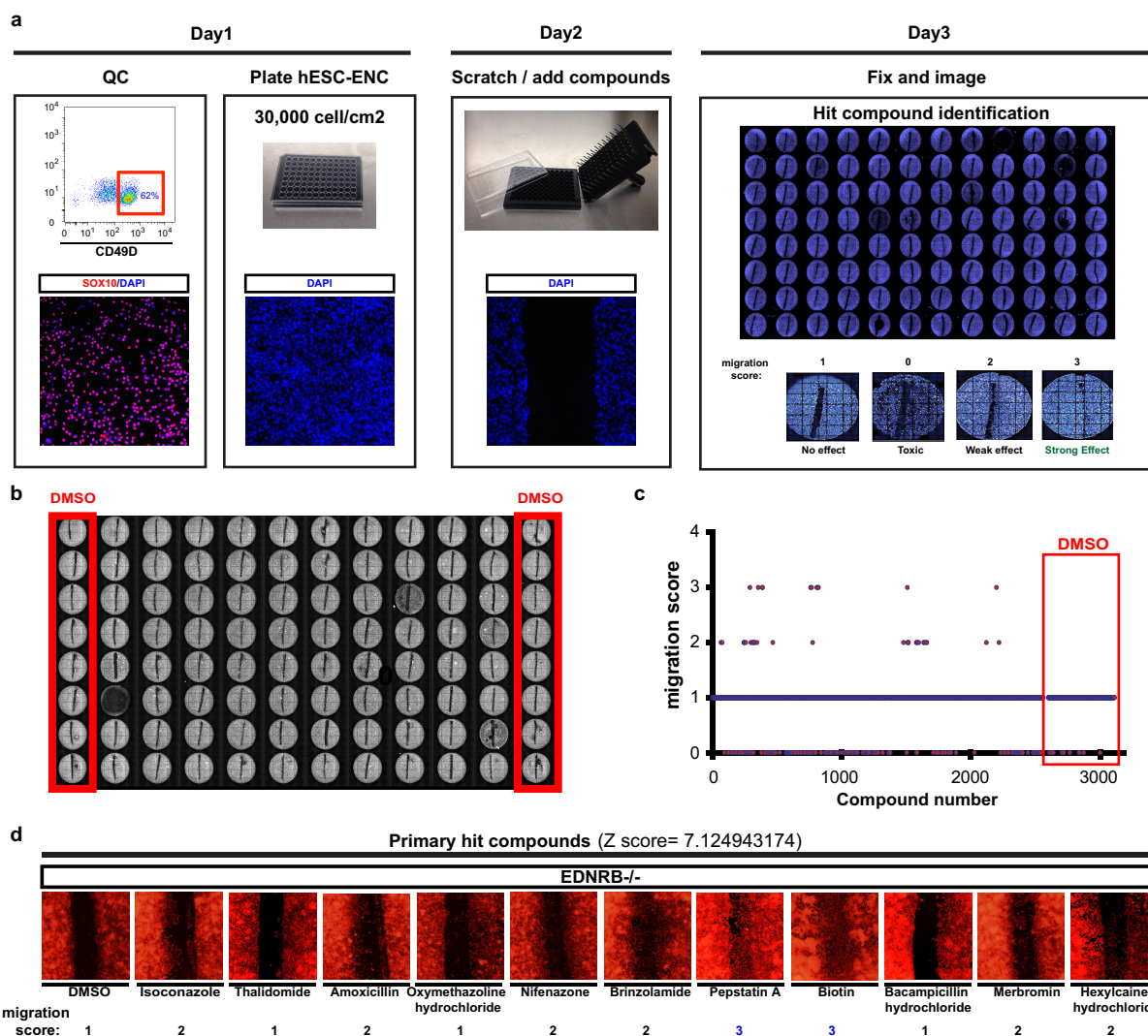
**Extended Data Figure 7 | Establishing and characterizing EDNRB-null human ES cell lines.** **a**, Sequences of wild-type and Cas9-nickase induced bi-allelic nonsense mutations in targeted region of *Ednrb*<sup>-/-</sup> clones.

**b**, Western blot analysis for EDNRB in ES-cell-derived ENC precursors showing lack of protein expression in the mutant clones C1–C4.

**c**, *EDNRB*<sup>-/-</sup> human ES cells can be efficiently differentiated into CD49D<sup>+</sup>

human ES-cell-derived ENC based on CD49D expression (**c**) and expression of SOX10 (data not shown). **d**, Proliferation of *EDNRB*<sup>-/-</sup> human ES-cell-derived ENCs (day 11) versus wild-type-derived cells; *n* = 4 independent experiments. **e**, LDH activity measurement of cell viability in *EDNRB*<sup>-/-</sup> ES-cell-derived ENCs (day 11) versus wild-type-derived cells.

\**P* < 0.05 (*t*-test; *n* = 3 independent experiments).

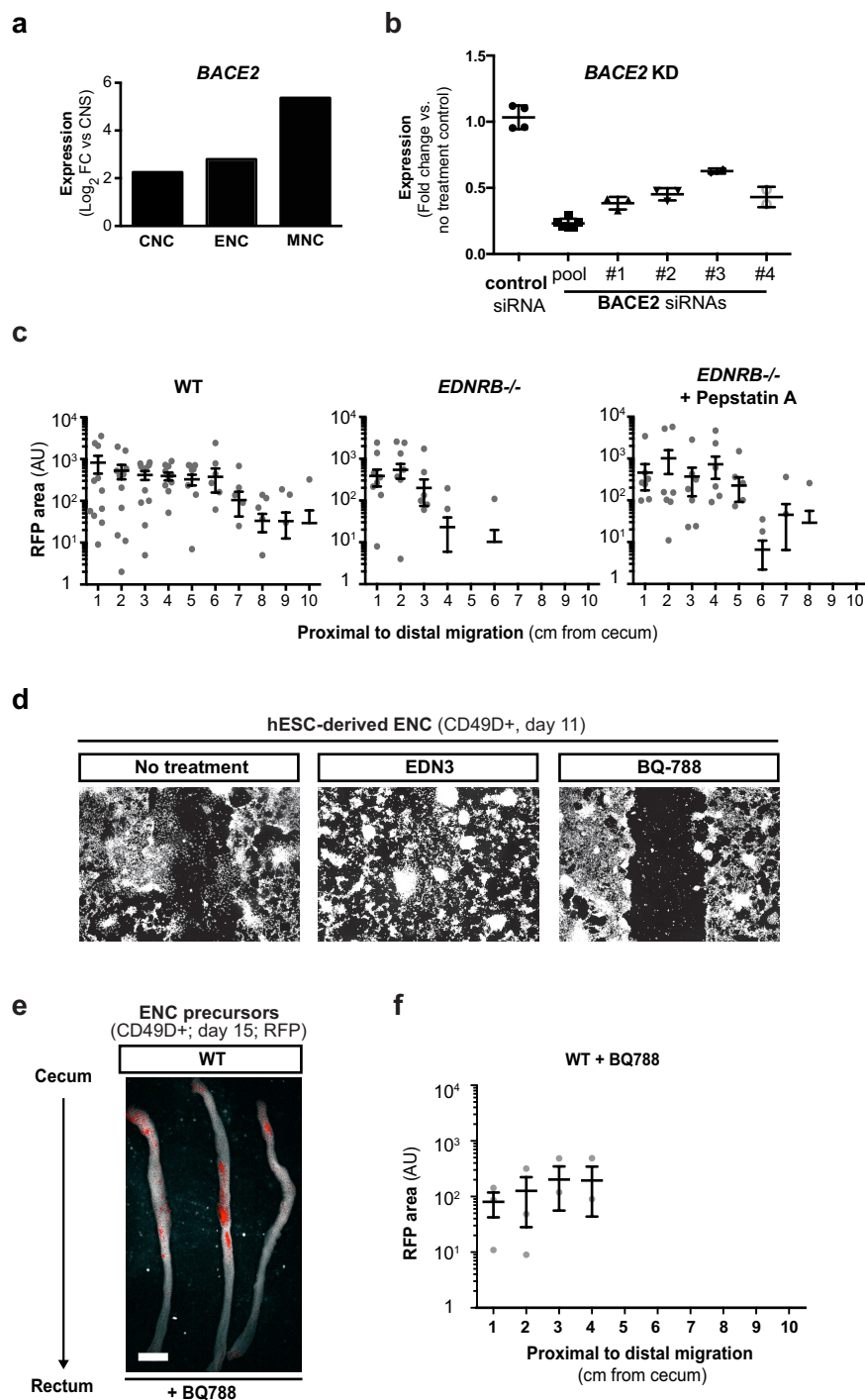


**Extended Data Figure 8 | Chemical screening for compounds that rescue migration of *EDNRB*<sup>-/-</sup> ES-cell-derived NC precursors.**

**a**, Schematic illustration of the timeline and experimental steps involved in the chemical screening assay and migration scoring system. **b**, Example of a screening plate layout and locations of dimethylsulfoxide (DMSO)

control wells. **c**, Migration scores of Prestwick library compounds and DMSO controls. **d**, Migration assay and scores for *EDNRB*<sup>-/-</sup> ES-cell-derived ENC precursors treated with primary hit compounds. Z-score for primary hit compounds in **c** is given numerically (compared to DMSO control;  $n = 224$  technical replicates).





**Extended Data Figure 9 | Pharmacological modulation of migration in human ES-cell-derived ENC precursors.** **a**, *BACE2* expression in the various human ES-cell-derived NC sublineages at day 11 as compared to stage-matched CNS precursors. **b**, qRT-PCR analysis to confirm knockdown of *BACE2* in CD49D-purified *EDNRB*<sup>-/-</sup> ES-cell-derived ENCs after siRNA transfection compared to control siRNA. **c**, Quantification of whole-mount images of the colon of NSG mice transplanted with RFP<sup>+</sup>

CD49D-purified wild-type and *Ednrb*<sup>-/-</sup> ES-cell-derived ENC precursors, with or without pepstatin A pre-treatment (compare to Fig. 4j). **d**, Representative images of wild-type CD49D-purified ES-cell-derived ENCs treated with EDN3 and BQ-788 (*EDNRB* inhibitor). **e**, Colon migration assay in wild-type ES-cell-derived ENC precursors after pre-treatment with BQ-788. **f**, Quantification of the data in **e**. AU, arbitrary unit. Scale bar, 1 cm.

# Mutant *Kras* copy number defines metabolic reprogramming and therapeutic susceptibilities

Emma M. Kerr<sup>1</sup>, Edoardo Gaude<sup>1</sup>, Frances K. Turrell<sup>1</sup>, Christian Frezza<sup>1</sup> & Carla P. Martins<sup>1</sup>

The RAS/MAPK (mitogen-activated protein kinase) signalling pathway is frequently deregulated in non-small-cell lung cancer, often through KRAS activating mutations<sup>1–3</sup>. A single endogenous mutant *Kras* allele is sufficient to promote lung tumour formation in mice but malignant progression requires additional genetic alterations<sup>4–7</sup>. We recently showed that advanced lung tumours from *Kras*<sup>G12D/+</sup>;p53-null mice frequently exhibit *Kras*<sup>G12D</sup> allelic enrichment (*Kras*<sup>G12D</sup>/*Kras*<sup>wild-type</sup> > 1) (ref. 7), implying that mutant *Kras* copy gains are positively selected during progression. Here we show, through a comprehensive analysis of mutant *Kras* homozygous and heterozygous mouse embryonic fibroblasts and lung cancer cells, that these genotypes are phenotypically distinct. In particular, *Kras*<sup>G12D/G12D</sup> cells exhibit a glycolytic switch coupled to increased channelling of glucose-derived metabolites into the tricarboxylic acid cycle and glutathione biosynthesis, resulting in enhanced glutathione-mediated detoxification. This metabolic rewiring is recapitulated in mutant KRAS homozygous non-small-cell lung cancer cells and *in vivo*, in spontaneous advanced murine lung tumours (which display a high frequency of *Kras*<sup>G12D</sup> copy gain), but not in the corresponding early tumours (*Kras*<sup>G12D</sup> heterozygous). Finally, we demonstrate that mutant *Kras* copy gain creates unique metabolic dependences that can be exploited to selectively target these aggressive mutant *Kras* tumours. Our data demonstrate that mutant *Kras* lung tumours are not a single disease but rather a heterogeneous group comprising two classes of tumours with distinct metabolic profiles, prognosis and therapeutic susceptibility, which can be discriminated on the basis of their relative mutant allelic content. We also provide the first, to our knowledge, *in vivo* evidence of metabolic rewiring during lung cancer malignant progression.

The Ras pathway<sup>8</sup> is frequently upregulated during the malignant progression of mutant *Kras* tumours<sup>5,9</sup>, indicating that this transition requires further increased Ras activity. But how this activity may contribute to malignant progression remains unclear. We recently identified mutant *Kras* (*Kras*<sup>mut</sup>) copy gains in high-grade murine lung tumours<sup>7</sup> and mutant-specific gains have also been reported in non-small-cell lung cancer (NSCLC)<sup>2,10</sup>. We thus hypothesized that the gain of a second *Kras*<sup>mut</sup> copy affords additional oncogenic phenotypes to *Kras* heterozygous cells. To identify such potential gain-of-function phenotypes, we compared the acute impact of *Kras*<sup>G12D</sup>-endogenous allele<sup>11</sup> activation in heterozygous and homozygous *Kras*<sup>G12D</sup> mouse embryonic fibroblasts (MEFs). MEFs were generated on a p53-null background<sup>12</sup> (Extended Data Fig. 1a) to recapitulate the tumour genotype in which *Kras*<sup>G12D</sup> copy gains were identified<sup>7</sup> but, for simplicity, hereafter they will be termed *Kras*<sup>wild-type/wild-type</sup> (WT/WT), *Kras*<sup>G12D/WT</sup> and *Kras*<sup>G12D/G12D</sup>.

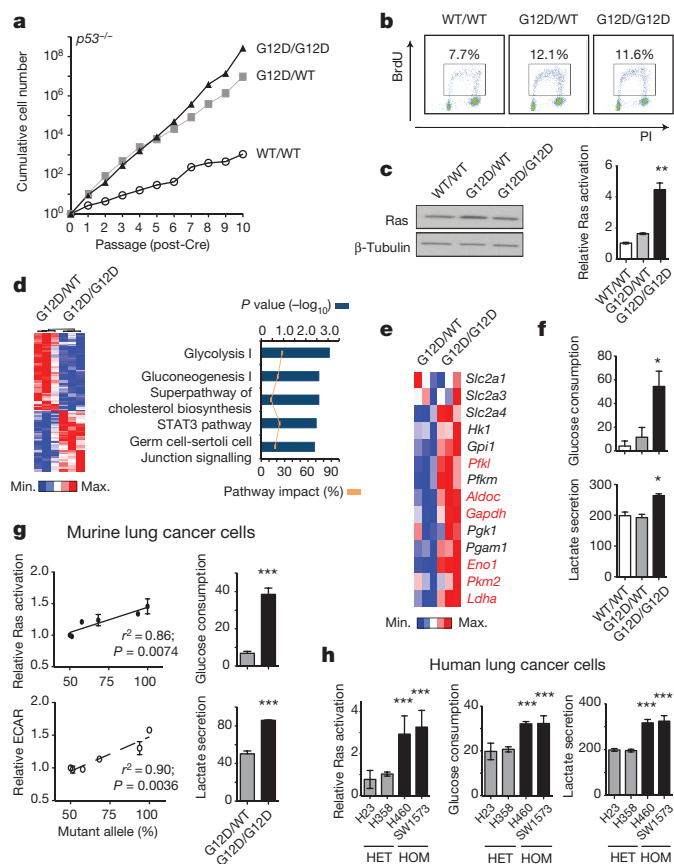
As reported<sup>11</sup>, *Kras*<sup>G12D/WT</sup> cells showed a proliferative advantage relative to *Kras*<sup>WT/WT</sup> MEFs. Surprisingly, *Kras*<sup>G12D/WT</sup> and *Kras*<sup>G12D/G12D</sup> cells grew similarly at early passages (P1–P5) (Fig. 1a, b), indicating that proliferation is not directly affected by *Kras*<sup>mut</sup> copy gain. A growth advantage of *Kras*<sup>G12D/G12D</sup> cells was nevertheless observed after P6.

To identify both immediate and proliferation-independent *Kras*<sup>G12D</sup> copy-gain-dependent effects, subsequent analyses were restricted to early passages. While KRAS amplifications are typically associated with increased expression<sup>2,10</sup>, *Kras*<sup>G12D/WT</sup> and *Kras*<sup>G12D/G12D</sup> Ras protein levels were comparable and only slightly increased relative to *Kras*<sup>WT/WT</sup> MEFs. Nevertheless, *Kras*<sup>G12D/G12D</sup> MEFs exhibited a ~2-fold increase in activated Ras relative to *Kras*<sup>G12D/WT</sup> cells (Fig. 1c), indicating that mutant copy gain may have functional implications. In agreement, microarray analysis identified 1,666 genes differentially regulated (>1.3-fold) between *Kras*<sup>G12D/G12D</sup> and *Kras*<sup>G12D/WT</sup> MEFs, with glycolysis being the most significantly altered pathway (Fig. 1d and Extended Data Fig. 1b).

Mutant *Kras* activity enhances glucose uptake and rewires glucose metabolism into the hexosamine biosynthesis and pentose phosphate pathways in pancreatic ductal adenocarcinoma<sup>13</sup>. However, its metabolic impact on other cancer types and, more importantly, that of *Kras*<sup>mut</sup> copy gain is unclear. *Kras*<sup>G12D/WT</sup> and *Kras*<sup>WT/WT</sup> MEFs showed similar glycolytic gene expression profiles with the exception of *Slc2a1* (*Glut1*) and *Slc2a3* (*Glut3*, data not shown). In contrast, in *Kras*<sup>G12D/G12D</sup> MEFs, glycolytic gene expression was significantly upregulated and mirrored by increased glucose uptake, lactate secretion and glycolytic capacity (Fig. 1e, f and Extended Data Fig. 1c, d). Thus, we show that *Kras*<sup>G12D</sup> copy gain induces a glycolytic switch while a *Kras* mutation per se is not sufficient to upregulate glycolysis. Notably, analysis of murine lung tumour cell lines with distinct *Kras* G12D/WT allelic content revealed a direct correlation between increased *Kras*<sup>G12D</sup> copy number (*Kras*<sup>G12D</sup>/total *Kras*) and enhanced GTPase activity as well as glycolysis (extracellular acidification rate, ECAR), consistent with a '*Kras*<sup>mut</sup>-dosage' effect. Glycolytic gene expression, glucose uptake and lactate secretion were also significantly enhanced in *Kras*<sup>G12D/G12D</sup> relative to heterozygous lung tumour cells (Fig. 1g and Extended Data Fig. 1e–g). Importantly, KRAS<sup>mut</sup> homozygosity is highly prevalent (48.6%) within mutant KRAS NSCLC cell lines (COSMIC), emphasizing its relevance and enabling the validation of our findings in a clinically relevant NSCLC model. Reassuringly, the distinct glycolytic phenotypes of KRAS<sup>mut</sup> heterozygous and homozygous cells were confirmed in NSCLC cells (Fig. 1h and Extended Data Table 1), demonstrating that glycolysis upregulation is a *Kras*<sup>mut</sup> copy-gain-associated gain-of-function.

Enhanced glycolysis is a well-recognized cancer phenotype typically associated with increased growth demands, and/or compensatory adaptation to mitochondrial defects<sup>14,15</sup>. Yet, early passage *Kras*<sup>G12D/G12D</sup> MEFs displayed a glycolytic switch relative to *Kras*<sup>G12D/WT</sup> cells despite exhibiting comparable proliferative rates, cell volume, diameter, protein and RNA content (Fig. 1a, b and Extended Data Fig. 2a–d). Furthermore, mitochondrial morphology and function were similar across genotypes (Extended Data Fig. 2e–h), despite a *Kras*<sup>G12D</sup>-associated decrease in membrane potential, as reported under overexpression conditions<sup>15,16</sup>. We then hypothesized that this glycolytic switch reflected alternative glucose utilization by *Kras*<sup>G12D/G12D</sup> cells. Liquid chromatography–mass spectrometry (LC–MS)-based metabolomics analysis confirmed the

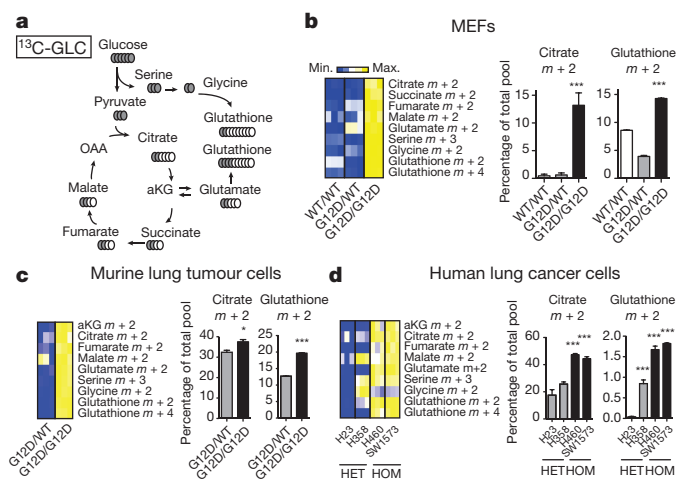
<sup>1</sup>MRC Cancer Unit, University of Cambridge, Box 197, Cambridge Biomedical Campus, Cambridge CB2 0XZ, UK.



**Figure 1 | Mutant *Kras* copy gain upregulates glycolysis in MEFs and lung tumour cells.** **a**, Proliferative rate of *Kras*<sup>WT/WT</sup> (WT/WT), *Kras*<sup>G12D/WT</sup> (G12D/WT) and *Kras*<sup>G12D/G12D</sup> (G12D/G12D); *p53*<sup>Fx/Fx</sup> MEFs. *p53* is also known as *Trp53* in mice (*TP53* in humans). **b**, Fluorescence-activated cell sorting (FACS) analysis denoting BrdU+ MEFs. **c**, MEF Ras levels (immunoblotting) and activation (Raf-GST pull-down, normalized to WT/WT). **d**, Heatmap illustrating differential gene expression between G12D/WT and G12D/G12D MEFs (*n* = 3 per genotype, microarray); top canonical pathways altered shown (Ingenuity Pathway Analysis, IPA). **e**, Glycolytic gene expression (MEF microarray-based heatmap). Genes significantly upregulated in G12D/G12D cells highlighted (bold red, *t*-test). **f**, MEF glucose consumption and lactate secretion. **g**, Left: *Kras*<sup>G12D</sup>/*Kras*<sup>Total</sup> allelic frequency (pyrosequencing) versus Ras activation or glycolysis (ECAR) in *Kras*<sup>G12D/WT</sup>; *p53*-deficient murine lung tumour cells (*n* = 6) (Pearson's correlation). Right: glucose consumption and lactate secretion in G12D/WT and G12D/G12D cell line pair (*t*-test). **h**, Ras activation (normalized to H358), glucose consumption and lactate secretion in *KRAS*<sup>mut</sup> heterozygous (HET: H23, H358) or homozygous (HOM: H460, SW1573) NSCLC cell lines. **c**, **f**, **h**, One-way analysis of variance (ANOVA). **a**–**c**, Representative data of three independent MEFs per genotype; **d**–**f**, *n* = 3 per genotype. **g**, **h**, Histograms: representative data (*n* = 3 independent experiments). All graphs depict triplicate mean ± s.d. (error bars). \*\*\**P* < 0.001; \*\**P* < 0.01; \**P* < 0.05.

enhanced glycolytic phenotype of *Kras*<sup>G12D/G12D</sup> cells and, unexpectedly, uncovered a significant increase in glucose-derived tricarboxylic acid (TCA) cycle metabolites in *Kras*<sup>mut</sup> homozygous MEFs and (murine and human) lung tumour cells (Fig. 2 and Extended Data Figs 3, 4), confirming their intact mitochondrial function. More importantly, these data identified a *Kras*<sup>mut</sup> copy-gain-specific metabolic rewiring and uncovered a (TCA-coupled) glucose metabolism signature not previously associated with mutant *Kras* activity.

Despite their differential glucose utilization, *Kras*<sup>G12D/WT</sup> and *Kras*<sup>G12D/G12D</sup> MEFs had similar oxidative phosphorylation levels (Extended Data Fig. 2e), hinting at additional TCA cycle differences. Since *Kras*<sup>mut</sup> cells were reported to preferentially utilize glutamine, rather than glucose, to fuel the TCA cycle<sup>17,18</sup>, glutamine metabolism was assessed. *Kras*<sup>G12D/WT</sup> cells had the highest levels of

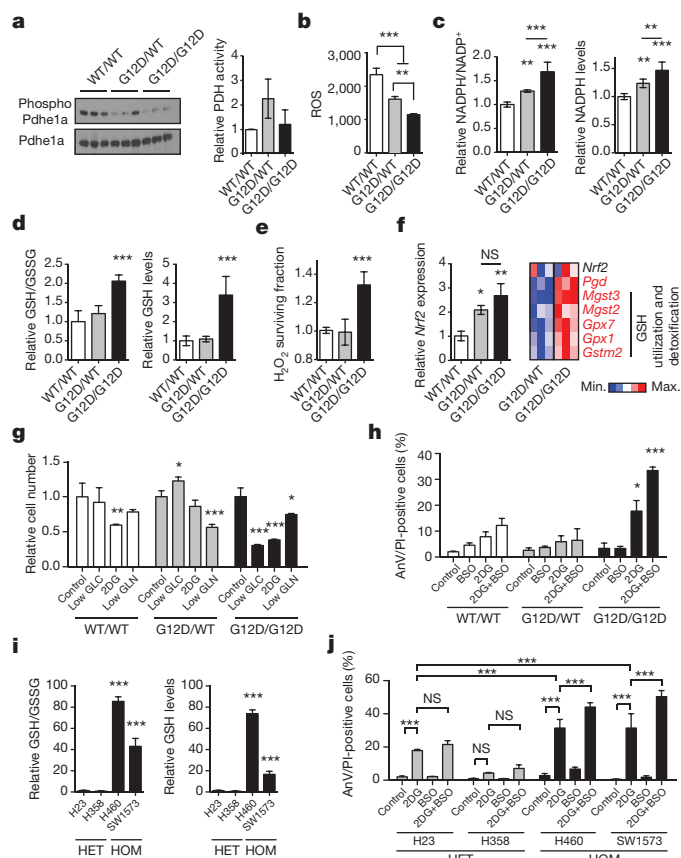


**Figure 2 | Mutant *Kras* copy gain drives glycolysis and directs glucose metabolism towards TCA cycle and glutathione synthesis.** Glucose metabolism flux analysis. **a**, Carbon flux (grey) from uniformly labelled <sup>13</sup>C-glucose (<sup>13</sup>C-GLC) illustrated. Glucose metabolism profiles (calculated as a percentage of the total metabolite pool) of indicated MEFs (**b**), murine (**c**) and human lung cancer cells (**d**) following LC-MS analysis. **b**–**d**, Representative data depict abundance of selected labelled metabolites. Triplicates (heatmaps) and triplicate mean ± s.d. (graphs) shown. MEFs and human cell lines: one-way ANOVA; murine cell lines: *t*-test. \*\*\**P* < 0.001; \**P* < 0.05.

glutamine-derived TCA cycle metabolites and glutamine-derived oxygen consumption was enhanced in *Kras*<sup>G12D/WT</sup>, but not *Kras*<sup>G12D/G12D</sup>, relative to *Kras*<sup>WT/WT</sup> MEFs (Extended Data Fig. 5a–j). However, unlike the genotype-specific glucose metabolism signatures, differential glutamine utilization could not be consistently recapitulated in tumour cells (data not shown), possibly reflecting their proliferative and oxygen consumption rate heterogeneity. Thus, *Kras*<sup>G12D/WT</sup>-specific glutamine metabolism rewiring is either MEF-specific or masked by other mutations in cancer cells.

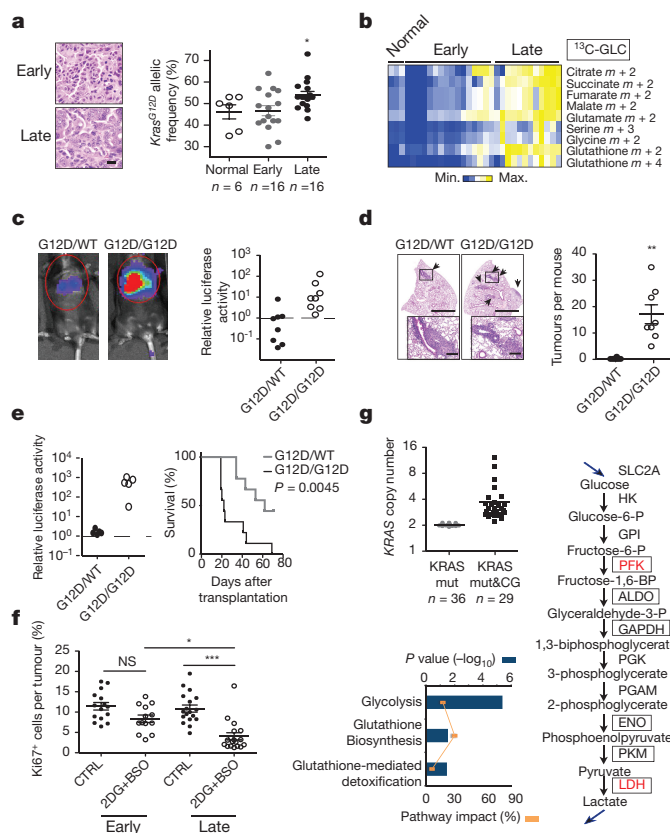
Enhanced pyruvate dehydrogenase (PDH) activity<sup>19</sup> could explain the glucose metabolism reprogramming exhibited by homozygous cells. However, since *Kras*<sup>G12D/G12D</sup> and *Kras*<sup>G12D/WT</sup> MEFs showed comparable *Pdhe1a* expression and PDH activity (Fig. 3a), we speculated that, instead, genotype-specific metabolic requirements drive this metabolic switch. Our metabolomics data showed that glutathione (GSH) and its precursors serine, glycine and glutamate were strikingly enriched with glucose-derived carbons in *Kras*<sup>mut</sup> homozygous cells (Fig. 2 and Extended Data Figs 3c–e, i and 4d–f, h, l, m), implying that *Kras*<sup>mut</sup> copy gain rewired glucose metabolism towards glutathione biosynthesis. Glutamine was also more efficiently metabolized towards GSH biosynthesis in *Kras*<sup>G12D/G12D</sup> MEFs (Extended Data Fig. 5g, k). We therefore assessed the impact of *Kras*<sup>G12D</sup>-gain on redox management. Consistent with previous reports<sup>18,20</sup>, *Kras*<sup>G12D/WT</sup> MEFs showed decreased ROS levels and an increased NADPH/NADP<sup>+</sup> ratio relative to *Kras*<sup>WT/WT</sup>. Nevertheless, *Kras*<sup>G12D/G12D</sup> MEFs exhibited a more striking antioxidant signature, marked by significantly increased NADPH and GSH synthesis, NADPH/NADP<sup>+</sup> and GSH/GSSG ratios and, conversely, lower ROS levels and increased resistance to ROS-inducing agents (H<sub>2</sub>O<sub>2</sub>) (Fig. 3b–e and Extended Data Fig. 6a). Mutant *Kras* was previously associated with enhanced expression and activity of the antioxidant programme regulator *Nrf2* (ref. 21) in *Kras*<sup>G12D/WT</sup>; *p53*<sup>+/+</sup> MEFs<sup>20</sup>, potentially explaining the increased redox potential of our *Kras*<sup>G12D/G12D</sup>; *p53*<sup>−/−</sup> MEFs. Despite exhibiting comparable *Nrf2* expression to heterozygous (Fig. 3f), *Kras*<sup>G12D/G12D</sup> MEFs showed upregulation of *Nrf2*-regulated GSH utilization genes<sup>21</sup>, indicating that increased *Nrf2*-mediated detoxification may contribute to their metabolic rewiring.





**Figure 3 | Mutant *Kras* copy-number dictates redox state, metabolic dependencies and therapeutic susceptibilities.** **a**, Total and phosphorylated Pdh1a levels and PDH activity in MEFs. **b**, Cellular ROS (CellRox); **c**, NADPH/NADP<sup>+</sup> ratio and NADPH levels; **d**, GSH/GSSG ratio and GSH levels in MEFs. **e**, MEF survival upon 24 h H<sub>2</sub>O<sub>2</sub> treatment. **f**, *Nrf2* and *Nrf2*-target gene expression in MEFs (left: quantitative PCR (qPCR); right: microarray). *Nrf2* targets significantly upregulated in homozygous MEFs highlighted (bold red, *t*-test). **g**, MEF viability after 72 h culture in low glucose (GLC), 2DG or low glutamine (GLN), relative to normal media (control). **h**, Percentage of AnnexinV<sup>+</sup>/propidium iodide<sup>+</sup> (AnV/PI)-positive, FACS (MFI) MEFs upon 48 h BSO, 2DG, or combined (2DG + BSO) treatment. **i**, GSH/GSSG ratio and GSH levels in *KRAS*<sup>mut</sup> NSCLC cells. **j**, NSCLC cells treated as in **h**. Triplicate mean  $\pm$  s.d. shown for three independent MEFs per genotype (**a**, **c**, **d**, **f**) or for representative data (three independent runs) (**b**, **e**, **g**–**j**). Data normalized to WT/WT (**a**, **c**–**f**) or HET mean (**i**). One-way (**a**–**f**, **i**) or two-way ANOVA (**g**, **h**, **j**). \*\*\**P* < 0.001; \*\**P* < 0.01; \**P* < 0.05; NS, not significant.

The metabolic heterogeneity of *Kras*<sup>mut</sup> cells can potentially limit the efficacy of generalized targeting approaches<sup>22</sup>, prompting us to explore potential *Kras*<sup>mut</sup> copy number-dependent susceptibilities. Unlike heterozygotes, *Kras*<sup>G12D/G12D</sup> MEFs were very sensitive to low glucose levels and the glucose analogue 2-deoxy-D-glucose (2DG), which induced a notable apoptotic response (Fig. 3g, h). In turn, *Kras*<sup>G12D/WT</sup> MEFs showed higher sensitivity to low glutamine. Confirming a reliance on glucose for efficient ROS management, *Kras*<sup>G12D/G12D</sup> cells (but not *Kras*<sup>WT/WT</sup> or *Kras*<sup>G12D/WT</sup>) showed increased ROS levels upon 2DG treatment, while *N*-acetyl-L-cysteine (NAC, GSH precursor) partly rescued their 2DG-induced apoptosis (Extended Data Fig. 6b, c). Moreover, combined 2DG and L-buthionine-S,R-sulfoximine (BSO, GSH biosynthesis inhibitor<sup>23</sup>) treatment induced drastic, *Kras*<sup>G12D/G12D</sup>-specific apoptosis and a reduction in GSH/GSSG ratio (Fig. 3h and Extended Data Fig. 6d). Likewise, murine and human *Kras*<sup>mut</sup> homozygous lung cancer cells exhibited increased GSH levels, GSH/GSSG ratio and enhanced sensitivity to low glucose, 2DG and 2DG/BSO, relative to heterozygotes (Fig. 3i, j and Extended Data Fig. 6e–g),



**Figure 4 | Mutant *Kras* copy gain results in increased malignancy and metabolic rewiring *in vivo*.** **a**, Representative haematoxylin and eosin (H&E) sections (scale bar, 20  $\mu$ m) and *Kras*<sup>G12D</sup> allelic frequency (pyrosequencing) in independent *Kras*<sup>G12D/+</sup>; *p53*<sup>Fx/Fx</sup> early and late lung tumours (*n* = 4 mice per cohort) and normal lung ('early', 'late', 'normal', respectively). **b**, Relative abundance of selected <sup>13</sup>C-glucose-derived metabolites (LC–MS) in samples from **a** (*n* = 3 normal, *n* = 16 early, *n* = 12 late). **c**, Representative imaging and luciferase activity/mouse 3 weeks after MEF transplantation (*n* = 8 per genotype). **d**, Representative H&E and quantification of lung tumours in MEF recipients (*t*-test). Arrows, lung tumours; scale bars, 2 mm (large), 250  $\mu$ m (small). **e**, Left: luciferase imaging of lung cancer cell (L1212, L1211) recipients, 3 weeks after transplantation (*n* = 5 per genotype, left). Right: recipient survival (Kaplan–Meier log-rank test, *n* = 9 per genotype). **f**, Ki67<sup>+</sup> quantification of early and late tumours treated for 48 h with 2DG + BSO or vehicle (CTRL) (*n* = 3 mice per cohort). **g**, *KRAS*<sup>mut</sup> TCGA lung adenocarcinoma<sup>1</sup> analysis following tumour segregation into 'KRAS<sup>mut</sup>' (mutation (mut)) and 'KRAS<sup>mut&CG</sup>' (mutation and copy gain (mut&CG)) cohorts. *KRAS* copy number/tumour shown (upper left). Differential expression of glycolysis and glutathione pathway genes illustrated (RNaseq, IPA; bottom left). Glycolytic genes significantly upregulated in *KRAS*<sup>mut&CG</sup> tumours (bold red, RNaseq) or G12D/G12D MEFs (boxes, microarray) relative to heterozygous illustrated (right). Mean  $\pm$  s.e.m. (**a**, **f**) or  $\pm$  s.d. (**d**) shown. **a**, **e**, One-way ANOVA. \*\*\**P* < 0.001; \*\**P* < 0.01; \**P* < 0.05; NS, not significant.

revealing a mutant *Kras* copy-gain-specific susceptibility to glucose and glutathione depletion in lung cancer cells.

Finally, we defined the metabolic impact of *Kras*<sup>G12D</sup> copy gain *in vivo* using the spontaneous *Kras*<sup>G12D/+</sup>; *p53*<sup>−/−</sup> lung tumours where these gains were originally reported<sup>7</sup>. These tumours progress over time from low- to high-grade, with *Kras*<sup>G12D</sup> gains being associated with the latter, prompting us to compare glucose flux in early (mostly low-grade) and late (typically advanced) tumours. Control or tumour-bearing mice were infused with <sup>13</sup>C-glucose and normal lung or individual tumours isolated for LC–MS analysis and biopsied for *Kras* locus assessment. Early tumours and control lung showed similar *Kras*<sup>G12D</sup> allelic content (means 46.7% and 46.2% respectively; Fig. 4a), demonstrating that early lesions are predominantly

heterozygous. In contrast, late tumours showed increased *Kras*<sup>G12D</sup> allelic prevalence (mean > 50%), confirming mutant enrichment in advanced disease<sup>7</sup>. Importantly, and consistent with our *in vitro* data, late tumours exhibited an increase in glucose-derived TCA cycle metabolites, as well as serine, glycine and GSH (Fig. 4b and Extended Data Fig. 7). Notably, by showing that early and late tumours have distinct metabolic profiles, we provide, to our knowledge, the first evidence of *in vivo* metabolic reprogramming during lung cancer malignant progression.

*Kras*<sup>G12D</sup> copy gain also drove increased malignancy in MEFs and lung cancer cells (Fig. 4c–e). Accordingly, *Kras*<sup>G12D/G12D</sup> MEFs showed a highly penetrant colonization phenotype inducing lung tumours in eight out of eight recipient mice following intravenous transplantation. In contrast, none of the *Kras*<sup>WT/WT</sup> (*n* = 5, data not shown) and only one out of eight *Kras*<sup>G12D/WT</sup> recipients developed a lung lesion. Similarly, *Kras*<sup>G12D/G12D</sup> lung cancer cells exhibited a significantly increased metastatic potential relative to heterozygous cells, establishing a direct link between *Kras*<sup>mut</sup> gains and lung cancer malignancy.

Lastly, we defined the therapeutic impact of glucose metabolism rewiring *in vivo* by treating early and late lung tumours with 2DG + BSO. Similarly to MEFs and lung cancer cells, late tumours (where *Kras*<sup>G12D</sup> allele prevalence is increased (Fig. 4a)) were significantly more sensitive than early tumours to 2DG + BSO treatment (Fig. 4f). Thus, despite the presence of a *Kras*<sup>G12D</sup> allele (and p53 inactivation) in both groups and their comparable proliferation (Fig. 4f, CTRL), low and high-grade lung tumours have distinct and mutant *Kras* copy number-dependent therapeutic susceptibilities.

*Kras* allelic imbalance<sup>7</sup> and mutant *Kras* upregulation<sup>9</sup> were shown to select for p53 inactivation and correlate with tumour progression, but the oncogenic effects of enhanced mutant *Kras* signalling remained unclear. Here we show that even in the absence of p53, *Kras*<sup>G12D/WT</sup> and *Kras*<sup>G12D/G12D</sup> cells are phenotypically distinct, with mutant *Kras* copy gain driving gain-of-functions that include upregulation and reprogramming of glucose metabolism, enhanced ROS management and increased metastatic potential. It is possible that loss of the *Kras*<sup>WT</sup> allele contributes to the phenotypes observed in *Kras*<sup>G12D/G12D</sup> cells<sup>24,25</sup>. However, since the majority of advanced murine lung tumours retain the wild-type allele (Fig. 4a), we argue that mutant *Kras* gains are the more likely target of positive selection during lung cancer progression. In agreement, *KRAS*<sup>WT</sup>-loss is uncommon whereas mutant and wild-type *KRAS* gains are frequent features of mutant *KRAS* human lung adenocarcinoma<sup>2,10,26</sup>.

Importantly and consistent with our findings, TCGA copy number variation and RNaseq data analysis of 65 mutant *KRAS* lung adenocarcinomas<sup>1</sup> revealed that, despite likely allelic gain heterogeneity<sup>10</sup>, combined *KRAS* mutation and copy gain correlate with glycolysis and glutathione metabolism pathway upregulation (*KRAS*<sup>mut&CG</sup>, Fig. 4g and Extended Data Table 2). These data confirm that mutant *KRAS* lung tumours are not a single metabolic entity and that, similarly to murine tumours, they may comprise (at least) two disease subgroups with distinct genetic and metabolic signatures and unique therapeutic susceptibilities. We argue that this heterogeneity may have contributed to the poor treatment responses of *KRAS* mutant tumours and hence that combined quantitative and qualitative *KRAS* locus assessment may have both prognostic and therapeutic utility.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 October 2015; accepted 5 January 2016.

Published online 24 February 2016.

1. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
2. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).

3. Vicent, S. *et al.* ERK1/2 is activated in non-small-cell lung cancer and associated with advanced tumours. *Br. J. Cancer* **90**, 1047–1052 (2004).
4. Guerra, C. *et al.* Tumor induction by an endogenous K-ras oncogene is highly dependent on cellular context. *Cancer Cell* **4**, 111–120 (2003).
5. Jackson, E. L. *et al.* The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65**, 10280–10288 (2005).
6. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243–3248 (2001).
7. Junttila, M. R. *et al.* Selective activation of p53-mediated tumour suppression in high-grade tumours. *Nature* **468**, 567–571 (2010).
8. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. *Nature Rev. Cancer* **11**, 761–774 (2011).
9. Sarkisian, C. J. *et al.* Dose-dependent oncogene-induced senescence *in vivo* and its evasion during mammary tumorigenesis. *Nature Cell Biol.* **9**, 493–505 (2007).
10. Modrek, B. *et al.* Oncogenic activating mutations are associated with local copy gain. *Mol. Cancer Res.* **7**, 1244–1252 (2009).
11. Tuveson, D. A. *et al.* Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. *Cancer Cell* **5**, 375–387 (2004).
12. Jonkers, J. *et al.* Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nature Genet.* **29**, 418–425 (2001).
13. Ying, H. *et al.* Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. *Cell* **149**, 656–670 (2012).
14. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
15. Hu, Y. *et al.* K-ras<sup>G12V</sup> transformation leads to mitochondrial dysfunction and a metabolic switch from oxidative phosphorylation to glycolysis. *Cell Res.* **22**, 399–412 (2012).
16. Baracca, A. *et al.* Mitochondrial complex I decrease is responsible for bioenergetic dysfunction in K-ras transformed cells. *Biochim. Biophys. Acta* **1797**, 314–323 (2010).
17. Gaglio, D. *et al.* Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth. *Mol. Syst. Biol.* **7**, 523 (2011).
18. Son, J. *et al.* Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature* **496**, 101–105 (2013).
19. Kaplan, J. *et al.* A key role for mitochondrial gatekeeper pyruvate dehydrogenase in oncogene-induced senescence. *Nature* **498**, 109–112 (2013).
20. DeNicola, G. M. *et al.* Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature* **475**, 106–109 (2011).
21. Gorrini, C., Harris, I. S. & Mak, T. W. Modulation of oxidative stress as an anticancer strategy. *Nature Rev. Drug Discov.* **12**, 931–947 (2013).
22. Bryant, K. L., Mancias, J. D., Kimmelman, A. C. & Der, C. J. KRAS: feeding pancreatic cancer proliferation. *Trends Biochem. Sci.* **39**, 91–100 (2014).
23. Griffith, O. W. Mechanism of action, metabolism, and toxicity of buthionine sulfoximine and its higher homologs, potent inhibitors of glutathione synthesis. *J. Biol. Chem.* **257**, 13704–13712 (1982).
24. To, M. D., Rosario, R. D., Westcott, P. M., Banta, K. L. & Balmain, A. Interactions between wild-type and mutant Ras genes in lung and skin carcinogenesis. *Oncogene* **32**, 4028–4033 (2013).
25. Zhang, Z. *et al.* Wildtype Kras2 can inhibit lung carcinogenesis in mice. *Nature Genet.* **29**, 25–33 (2001).
26. Wagner, P. L. *et al.* Frequency and clinicopathologic correlates of KRAS amplification in non-small cell lung carcinoma. *Lung Cancer* **74**, 118–123 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank T. Jacks (*Kras*<sup>LSL-G12D</sup>), A. Berns (*p53*<sup>flx</sup>) and the National Cancer Institute Mouse Repository for mice. We also thank S. Kleeman, P. Ogger and S. Costa for assistance with redox cell profiling, cell viability assays and LC-MS, respectively. We are very grateful to Cancer Research UK Cambridge Institute Biological Resources Unit staff for support with *in vivo* work and all the members of the Martins laboratory for comments and advice. This work was supported by the Medical Research Council.

**Author Contributions** C.P.M. and E.M.K. designed the study and E.M.K. performed all experiments with assistance from F.K.T. (*in vivo* work, qPCR, Seahorse) and E.G. (PDH activity). E.G. and C.F. contributed to the design of metabolomics experiments and performed LC-MS. All authors contributed to data analysis and discussion. C.P.M. and E.M.K. wrote the paper and all authors contributed to editing.

**Author Information** Microarray data have been deposited in the Gene Expression Omnibus under accession number GSE75871. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.P.M. ([c.martins@mrc-cu.cam.ac.uk](mailto:c.martins@mrc-cu.cam.ac.uk)).



## METHODS

**Mice, adenoviral infection and treatments.** Animals were maintained under SPF conditions and in compliance with UK Home Office regulations. *Kras*<sup>LSL-G12D</sup> (ref. 6) mice were crossbred to *p53*<sup>Fx/Fx</sup> (ref. 12) to obtain mixed background (C57Bl/6/129/Sv) *Kras*<sup>LSL-G12D/+</sup>; *p53*<sup>Fx/Fx</sup> (for spontaneous tumours and MEF generation) or *Kras*<sup>+/+</sup>; *p53*<sup>+/+</sup> (transplantation recipients) mice. Endogenous lung tumours were generated through intranasal administration of 8- to 10-week-old *Kras*<sup>LSL-G12D/+</sup>; *p53*<sup>Fx/Fx</sup> mice (termed *Kras*<sup>G12D/WT</sup>; *p53*<sup>null</sup>) with Cre-expressing adenovirus ( $5 \times 10^7$  plaque-forming units per mouse, University of Iowa Vector Core), as previously described<sup>7</sup>. For therapeutic studies, tumour-bearing mice were treated 12 (early group) or 16 weeks (late group) after Cre administration with a combination of 1,000 mg kg<sup>-1</sup> 2DG and 10 mmol kg<sup>-1</sup> BSO or vehicle (saline) once a day for 2 days (intraperitoneal). Lungs were collected 24 h after the last treatment and formalin fixed (of note: 2DG + BSO treatment was sometimes associated with a temporary decrease in motility in both control and tumour-bearing mice).

For transplantation studies, 8- to 12-week-old syngeneic wild-type mice were sublethally irradiated (4 Gy, caesium source) 6 h before tail vein injection with  $1 \times 10^5$  cells in 100 µl PBS. Baseline luminescence values were collected 24 h after transplantation and tumour growth monitored weekly by bioluminescence imaging after intraperitoneal injection with D-luciferin (150 mg/kg body weight) using an IVIS Spectrum Xenogen machine (Caliper Life Sciences). Relative luciferase activity corresponds to change from baseline at indicated time point, normalized to blank control (luciferase-negative animal). For tumour load analysis, lungs were collected 3 weeks after transplantation, whereas tumour survival represents the onset of moderate signs of disease. Two independent MEFs per genotype were used in transplantation studies (four recipients per MEF line). Lung cancer cell lines (L1211 and L1212) were each transplanted onto five (tumour load study) or nine recipient mice (survival). All studies involved animals of both sexes and no animals were excluded from the analysis. Cohort sizes were calculated on the basis of published data<sup>7</sup>, and pilot studies and animals were randomized on the basis of gender and age. Tumour analysis was performed blindly. No tumour exceeded the maximum size approved by the animal welfare committee/regulations.

**Generation and culture of MEFs and tumour cell lines.** For MEF generation, *Kras*<sup>LSL-G12D/+</sup>; *p53*<sup>Fx/Fx</sup> animals were interbred and embryos collected at embryonic day (E)12.5, to overcome *Kras*<sup>LSL-G12D/G12D</sup> embryonic lethality<sup>27</sup>, and Cre-mediated recombination performed immediately after MEF generation. In short, cells were cultured in DMEM supplemented with 10% FBS, 2 mM L-glutamine for one passage and then infected with adenovirus-Cre ( $5 \times 10^7$  plaque-forming units per  $1 \times 10^6$  cells). Cre-mediated recombination was confirmed by PCR. MEF data were typically obtained using three independent MEFs per genotype and a minimum of two independent MEFs per genotype was used in all cases. All (short-term) assays were performed in low-passage MEFs (P1–P4 post-Cre). For assessment of proliferative capacity, MEFs were cultured under standard 3T3 protocol. Briefly, at every passage  $3 \times 10^5$  cells were plated in triplicate on 6 cm plates and counted 3 days later. Cumulative cell number was calculated as  $\log(N_f/N_i)/\log_2$ , where  $N_i$  and  $N_f$  correspond to number of cells plated and final counts/passage, respectively. Murine lung tumour cell lines were generated from independent, spontaneous 'late' lung tumours from three *Kras*<sup>LSL-G12D/+</sup>; *p53*<sup>R270H/ER</sup> (refs 6, 28) mice. Tumour cells were dissociated by collagenase/dispase (Roche) treatment and cultured in DMEM/F12 media supplemented with 10% FBS, 2 mM L-glutamine. Human NSCLC cell lines were recently purchased from ATCC (authenticated) and cultured in RPMI media supplemented with 10% FBS, 2 mM L-glutamine.

AnnexinV/PI FACS analysis was performed as described<sup>29</sup>. For BrdU/PI FACS cells were labelled with 10 mM BrdU (Sigma) for 2 h. After harvest, cells were fixed and stained with FITC-conjugated Anti-BrdU monoclonal antibody (Becton Dickinson), according to the manufacturer's protocol and resuspended in PBS containing 20 µg ml<sup>-1</sup> of propidium iodide (PI). FACS was performed using a LSRII (BD) flow cytometer and analysed with FlowJo software (Treestar). Cell viability following nutrient deprivation and H<sub>2</sub>O<sub>2</sub> administration was determined by Trypan Blue exclusion (0.4%, Gibco) or Crystal Violet (0.2%, Sigma), respectively. For *in vivo* luciferase imaging, MEFs and tumour cells were transduced with an MSCV-luciferase-hygromycin retrovirus and selected (350 µg ml<sup>-1</sup> hygromycin B) before intravenous transplantation. All cells used in this study tested negative for mycoplasma contamination. *In vitro* assays were performed in triplicate and run at least three independent times.

**Ras and PDH expression and activity.** Ras activation was determined by Raf-GST pull-down-based ELISA using 100 µg protein/sample (whole-cell lysates) (Merck Millipore, Ras activation ELISA kit, 17-497). PDH activity was determined with a PDH enzyme activity assay kit (ab109902, Abcam), according to the manufacturer's instructions. Immunoblotting (40 µg protein per sample) was performed with anti-Ras (Cell Signalling, 3339), anti-phospho-Pdhe1a (Ser 293; AP1062; Calbiochem), total Pdhe1a (9H9AF5; 459400; Life Technologies) or anti-β-tubulin (Cell Signalling 2146, loading control) antibodies.

**Gene expression profiling, IPA and qPCR validation.** Microarray analysis was performed on three independent MEFs per genotype using Illumina MouseWG-6 version 2.0 Expression BeadChip (Department of Pathology, Cambridge University). Normalized log<sub>2</sub> values were determined and average log fold change (logFC) calculated for each comparison. Pathway analysis of genes differentially expressed (>1.3-fold) between genotypes was performed using IPA software (<http://www.ingenuity.com>) and statistical significance ( $P < 0.05$ ) of canonical pathways determined by Benjamini–Hochberg multiple testing correction. Relative gene expression was depicted by heatmaps generated using GENE-E software and statistical significance ( $P < 0.05$ ) determined by *t*-test. Gene expression changes were validated by qPCR using ROCHE Universal Probe Library System or Life Technologies probes and all data normalized to 18S expression.

**Immunohistochemistry.** H&E and Ki67 (Bethyl Laboratories, IHC-00375) immunohistochemistry was performed on formalin-fixed, 5 µm paraffin-embedded tissue sections. For transplantation studies, the total number of tumours in a single representative H&E section per animal (minimum of four lung lobes per section) is shown. For endogenous tumours, the percentage of proliferating cells was determined as Ki67<sup>+</sup>/DAPI<sup>+</sup> nuclei per tumour from a single representative section per animal (minimum of four lung lobes per section). All cells or a minimum of 4,000 (and >50% coverage) DAPI<sup>+</sup> nuclei per tumour were counted from an average of 5.25 tumours per mouse.

**Extracellular flux profiling.** Oxygen consumption rate and ECAR levels were determined using a Seahorse XF<sup>24</sup> analyser. Twenty thousand MEFs or  $4 \times 10^4$  tumour cells were plated in Seahorse XF<sup>24</sup> assay plates. Immediately before analysis, media was replaced by bicarbonate-free DMEM (Sigma) supplemented with 143 mM NaCl, 2% FBS and, where appropriate, 25 mM Glucose, 4 mM L-glutamine, pH 7.4 and cells incubated at 37 °C for 30 min in a CO<sub>2</sub>-free incubator. Each cycle of measurement involved 3 min mixing, 3 min waiting and 3 min measuring. After baseline measurements, testing agent prepared in assay medium was injected and followed by subsequent measuring cycles. Glycolysis stress test: measurement 1–3, basal (no glucose); 4–6, glucose (10 mM); 7–9, complex V inhibitor oligomycin (1 µM); and 10–12, 2-deoxyglucose (2DG, 100 mM). Mitochondrial stress test: measurement 1–3, normal (basal + 25 mM glucose); 4–6, oligomycin (1 µM); 7–9, carbonyl cyanide *m*-chlorophenylhydrazine (CCCP, 500 nM); and 10–12, rotenone (1 µM). Protein content (BCA assay, Thermo Fisher) at endpoint or cell number was used for data normalization.

**Metabolomics analysis.** *In vitro* sample preparation. Five hundred thousand MEFs or tumour cells were supplemented with media containing uniformly labelled <sup>13</sup>C-glucose (25 mM) or glutamine (4 mM) for 4 h before sampling. Metabolites were extracted from media (extracellular) and cell pellet (intracellular) in 50% MeOH: 30% Acetonitrile: 20% H<sub>2</sub>O + 100 ng ml<sup>-1</sup> HEPES buffer (1 ml per 10<sup>6</sup> cells). Samples were incubated at 4 °C for 15 min at 700 r.p.m., before centrifugation at 13,000 r.p.m. (15,700g). Supernatant was transferred to vials for mass spectrometry analysis.

*In vivo* sample preparation. *Kras*<sup>G12D/WT</sup>; *p53*<sup>Fx/Fx</sup> mice were anaesthetized (isoflurane inhalation) and administered a bolus of 0.4 mg g<sup>-1</sup> body weight <sup>13</sup>C-glucose by tail vein intravenous injection before continuous infusion of 0.012 mg g<sup>-1</sup> min<sup>-1</sup> at 150 µl h<sup>-1</sup> for 3 h. Normal lungs and independent lung tumours were collected and snap frozen. Samples were transferred to Precellys24 tubes, metabolite extraction buffer (as above) was added (250 µl per 10 mg), samples homogenized, centrifuged at 13,000 r.p.m. (15,700g) and supernatant used for mass spectrometry analysis. Before analysis, tissues were biopsied for *Kras* allelic assessment (pyrosequencing).

LC-MS metabolomics. Sequant Zic-pHilic (150 mm × 2.1 mm, internal diameter 5 µm) column and guard column (20 mm × 2.1 mm, internal diameter 5 µm) from HiChrom were used for LC separation. Mobile phase A: 20 mM ammonium carbonate plus 0.1% ammonia hydroxide in water. Mobile phase B: acetonitrile. Flow rate was maintained at 180 µl min<sup>-1</sup> and the gradient was as follows: 0–1 min 70% of B; 16 min 38% of B; 16.5 min 70% of B; 25 min 70% of B. A mass spectrometer (Thermo QExactive Orbitrap) was operated in full MS and polarity switching mode. Samples were randomized to avoid machine drifts and run in triplicate. Spectra were analysed using XCalibur Qual Browser and XCalibur Quan Browser softwares (Thermo Scientific) by referencing to an internal library of compounds. Relative metabolite abundance was calculated as the percentage of the indicated isotopologue to the total pool size of that metabolite, and depicted graphically or as a heatmap using GENE-E software. Samples were run and processed blindly.

**Glucose consumption and lactate measurements.** For glucose consumption analysis,  $1 \times 10^5$  MEF or tumour cells were incubated for 1 h with fluorescent 6-NBDG (N-23106, Life Technologies) and analysed by FACS (percentage of NBDG-positive cells). Lactate production was assessed 48 h after plating using Lactate Reagent (Trinity Biotech, Ireland), according to the manufacturer's instructions. Total lactate was normalized to cell number.



**TCGA data set analysis.** Lung adenocarcinoma patient TCGA data<sup>1</sup> (no restrictions) was downloaded from c-Biportal and KRAS mutation and copy number assessed. Tumours with a KRAS mutation and KRAS GISTIC score of 0+ were taken forward for analysis ( $n = 65$ ). Samples were divided into two cohorts based on GISTIC classification: KRAS<sup>mut</sup> (mutation only, GISTIC = 0,  $n = 36$ ) or KRAS<sup>mut&CG</sup> (mutation and copy gain, GISTIC = 1&2,  $n = 29$ ). KRAS copy number was calculated ( $2^{CNV} \times 2$ ) from SNP6 data. Pathway analysis was performed on the basis of RNaseq<sup>1</sup> metabolic gene expression data (1,430 genes) using IPA software, as described above.

**Reagents.** Metabolic probes: 6-NBDG (30  $\mu$ M), Mitotracker Green (50 nM), TMRM (50 nM), NAO (200 nM) and Cell Rox deep red (5  $\mu$ M) were obtained from Life Technologies. Cellular ROS was determined using CellRox deep red probe (Life Technologies, C10422) and FACS analysis. GSH/GSSG and NADPH/NADP levels and ratios were calculated according to the manufacturer's instructions (Promega, V6611 and G9081, respectively). Normal and low GLC and GLN (Sigma) correspond to 25 mM and 5 mM, and 4 mM and 0.5 mM, respectively. Cells were treated with the following agents, either alone or in combination as indicated: 100  $\mu$ M H<sub>2</sub>O<sub>2</sub>, 10 mM 2DG and 2 mM (tumour cells) or 5 mM BSO (MEFs). Total protein content was assessed using Sulforhodamine B (Sigma, 0.057% w/v), and total RNA content extracted with TRIzol (Life Technologies). NAC (4 mM, Sigma) was added to cells daily.

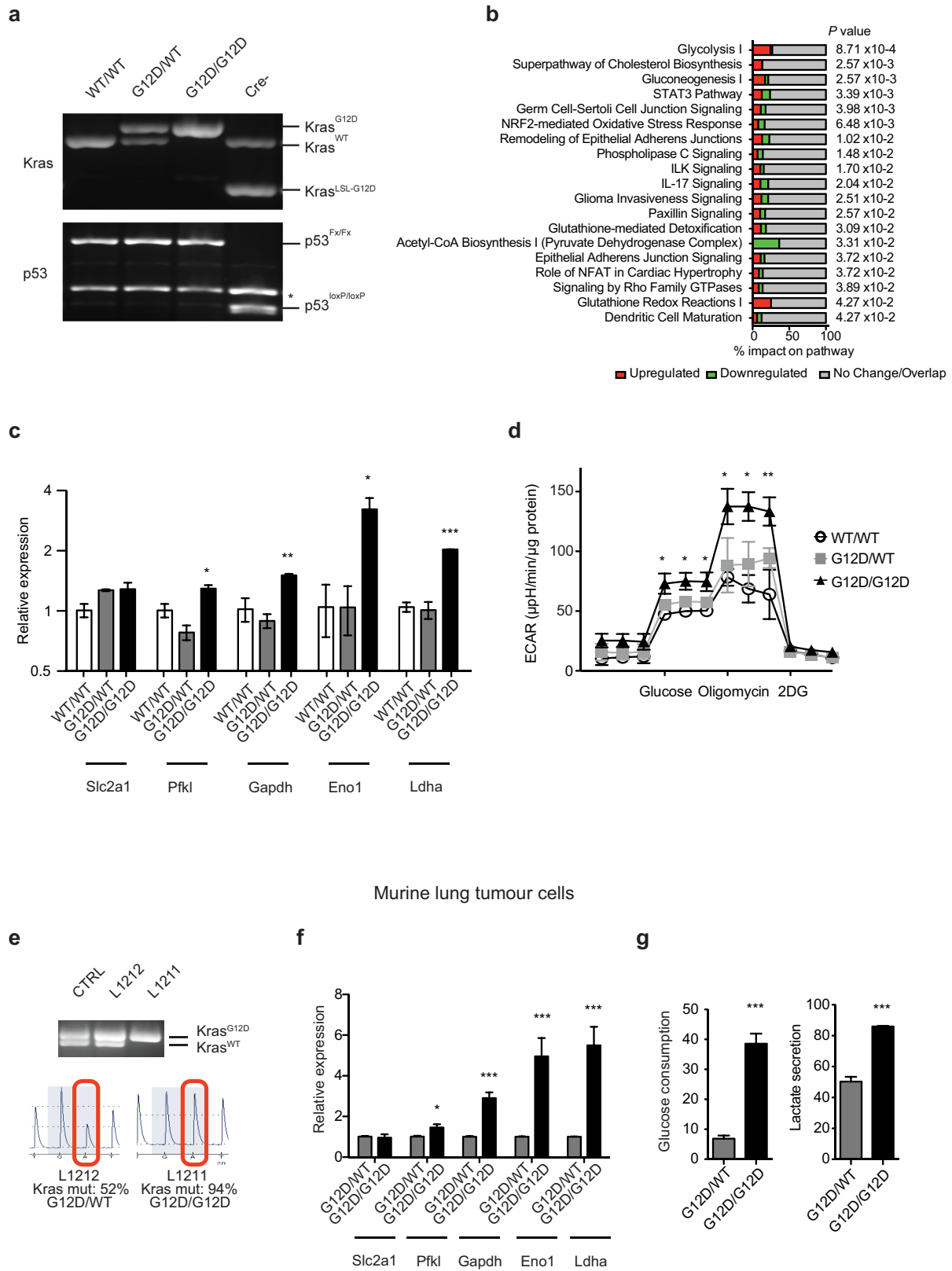
**Pyrosequencing analysis.** Genomic DNA (from MEFs, tumour cells and tumours) was isolated and the *Kras* WT/mutant allelic ratio determined by pyrosequencing (PyromarkQ24, Qiagen) according to the manufacturer's instructions. Pyrograms were analysed (PyroMarkQ24 software) and WT and mutant *Kras* allelic frequency determined on the basis of standards (WT:mutant ratios 2:0, 1:1, 0:2) and shown as the percentage of total *Kras* content.

**Primers and probes.** Genotyping: *Kras*<sup>LSL-G12D</sup> (ref. 6) and *p53*<sup>Fx</sup> (ref. 12) alleles were genotyped as reported. Microarray validation primers and probes were as follows. (UPL library, Roche): GAPDH: forward 5'-GGGTTCTTATAAATACGGACTGC-3', reverse 5'-CCATTTTGTCTACGGGACGA-3'; Probe 52;

Slc2a1: forward 5'-GGATCCCAGCAGCAAGAAG-3', reverse 5'-CCAGTGTATAGCCGAAGTGC-3'; Probe 76; Pfkf: forward 5'-GGGTCATGTACAGCGAGGA-3', reverse 5'-GGCCTCCATACCCATCTTG-3'; Probe 41; Eno1: forward 5'-GAGGACACTTTTCATCGCAGAC-3', reverse 5'-CCAGCTCTTCCTCAATTCTGA-3'; Probe 77; Nr1f2: Mm00477784\_m1, Ldha: Mm01612132\_m1 (Life Technologies), 18S: 4352930E (Thermo Fisher). Pyrosequencing: (Qiagen Q24 pyrosequencing assay) mKras<sup>G12D</sup> forward 5'-GTAAGGCCTGCTGAAAATGACTGA-3'; mKras<sup>G12D</sup> reverse 5'-[Btu]TATCGTCAAGGCGCTCTTGC-3', mKras<sup>G12D</sup> sequencing primer 5'-TGAAAATGACTGAGTATAAA-3'.

**Statistical analysis.** Data were visualized and statistical analyses performed using Prism 5.0 software (Graph Pad) or R statistical package.  $P < 0.05$  was considered statistically significant. In all cases, experimental groups showed comparable variance.  $P$  values for unpaired comparisons between two groups with comparable variance were calculated by two-tailed Student's  $t$ -test. One-way ANOVA (all groups against wild-type (WT) group) was used for analysis between three or more groups with comparable variance, followed by Bonferroni post-test for individual comparisons. Ordinary two-way ANOVA (treatment groups against individual genotype control group with comparable variance) was used for analysis that involved two variables, followed by Bonferroni post-test for individual comparisons. Kaplan–Meier comparison was used for analysis of survival cohorts. Pearson's correlation analysis was used to compare relationships between variables in groups with similar distribution. TCGA gene expression data were analysed using a negative binomial generalized linear model (DESeq2). \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . Error bars, mean  $\pm$  s.d. or s.e.m., as indicated.

27. Johnson, L. *et al.* K-ras is an essential gene in the mouse with partial functional overlap with N-ras. *Genes Dev.* **11**, 2468–2481 (1997).
28. Christophorou, M. A. *et al.* Temporal dissection of p53 function *in vitro* and *in vivo*. *Nature Genet.* **37**, 718–726 (2005).
29. Martins, C. P., Brown-Swigart, L. & Evan, G. I. Modeling the therapeutic efficacy of p53 restoration in tumors. *Cell* **127**, 1323–1334 (2006).

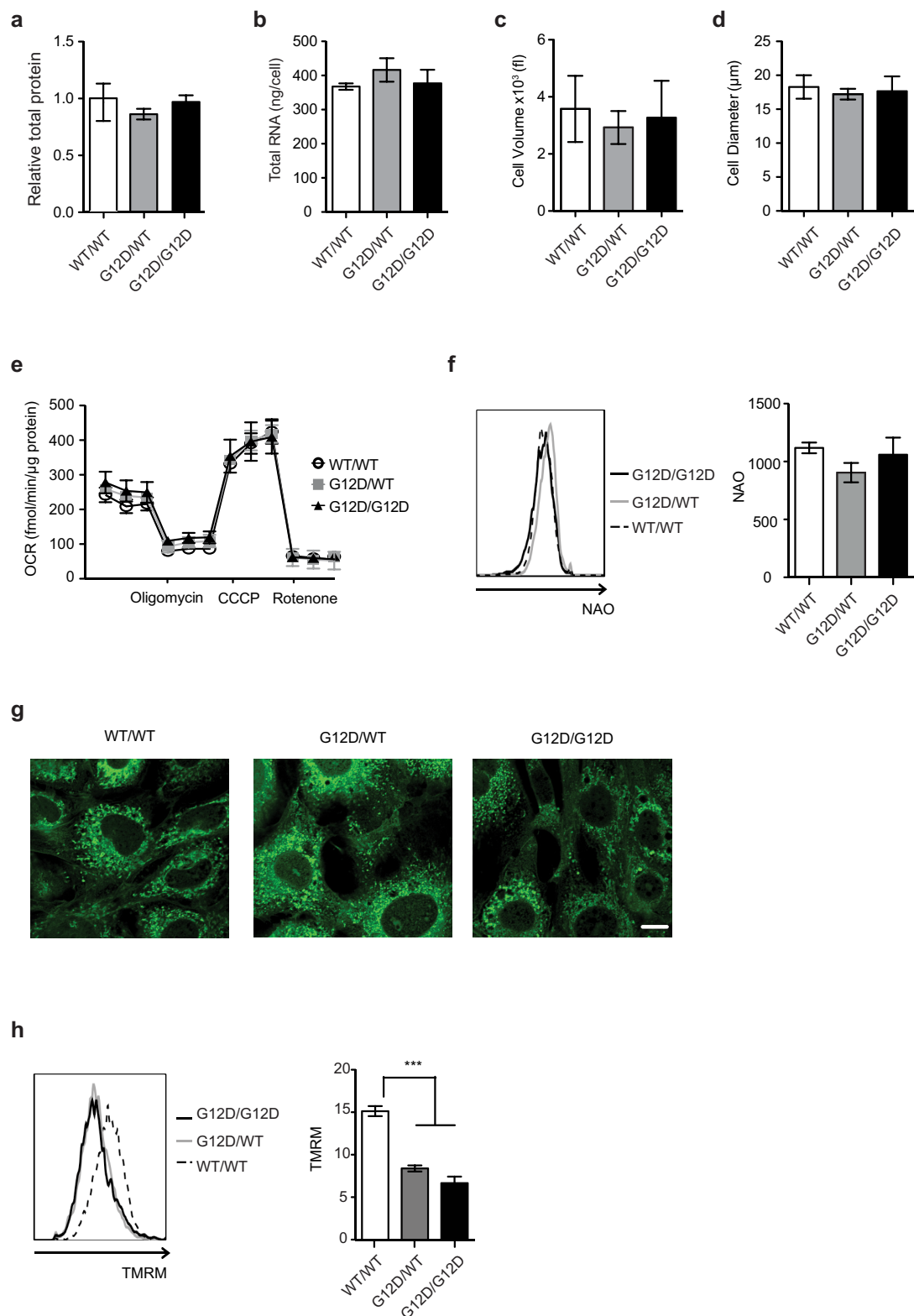


Extended Data Figure 1 | See next page for caption.

**Extended Data Figure 1 | Enhanced glycolysis in homozygous *Kras*<sup>G12D</sup> cells.** **a**, Representative data ( $n = 3$ ) of PCR analysis of the *Kras* and *p53* loci in *Kras*<sup>WT/WT</sup> (WT/WT), *Kras*<sup>G12D/WT</sup> (G12D/WT) and *Kras*<sup>G12D/G12D</sup> (G12D/G12D); *p53*<sup>Fx/Fx</sup> MEFs after Cre-mediated recombination; and of unrecombined *Kras*<sup>LSL-G12D/WT</sup>; *p53*<sup>loxP/loxP</sup> control (Cre-) (\*background band). **b**, IPA analysis of canonical pathways significantly altered in *Kras*<sup>G12D/G12D</sup> relative to *Kras*<sup>G12D/WT</sup> MEF transcriptomes ( $n = 3$  per genotype). **c**, Representative qPCR data ( $n = 3$ ) of glycolytic gene expression in MEFs. Fold change relative to WT/WT shown as triplicate mean  $\pm$  s.d. (one-way ANOVA). **d**, ECAR in MEFs following exposure to glucose, oligomycin and 2DG. Representative data from three independent MEFs per genotype show mean value between triplicates  $\pm$  s.d. (two-way

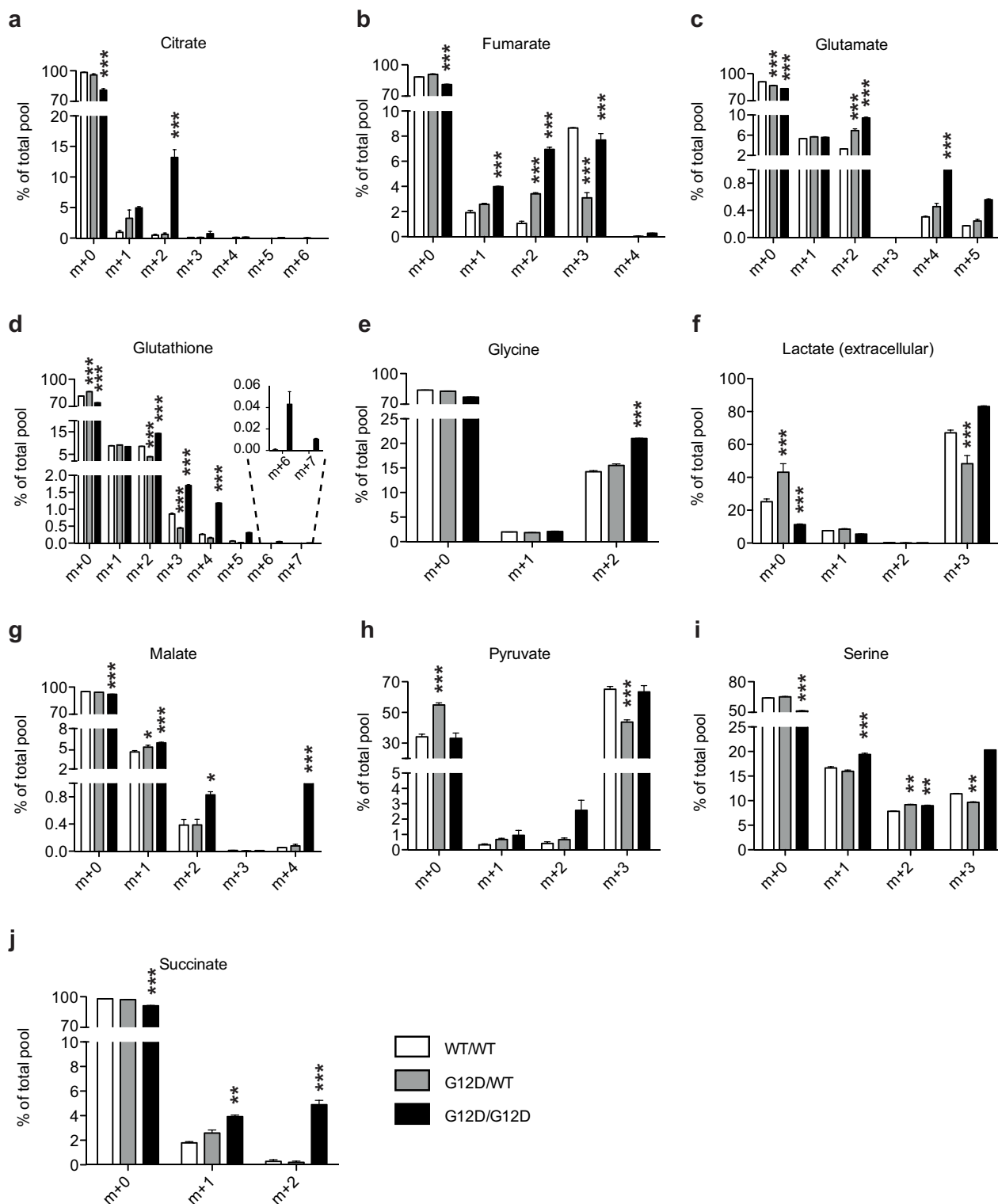
ANOVA). **e**, *Kras* locus analysis of two lung cancer cell lines (L1212 and L1211) generated from spontaneous tumours from *Kras*<sup>G12D/WT</sup>; *p53*-deficient mice. PCR (top) and pyrosequencing (bottom) analysis shown (L1212: *Kras* heterozygous, G12D/WT; L1211: G12D homozygous, G12D/G12D). Recombined heterozygous MEFs shown as PCR control (CTRL). **f**, Representative qPCR data ( $n = 3$ ) of glycolytic gene expression in L1211 and L1212 lung tumour cells. Fold change relative to heterozygous cells shown (mean of triplicates  $\pm$  s.d.; \*\*\* $P < 0.001$ ; \* $P < 0.05$ ;  $t$ -test). **g**, Left: basal glucose consumption in murine lung tumour cells determined by FACS analysis of 6-NBDG uptake (%). \* $P = 0.02$ ;  $t$ -test. Right: extracellular lactate concentration (ng/dl/cell) in murine lung tumour cells. Data are triplicate mean  $\pm$  s.d. \* $P = 0.0139$ ;  $t$ -test.





**Extended Data Figure 2 | *Kras*<sup>G12D/WT</sup> and *Kras*<sup>G12D/G12D</sup> MEFs have similar biomass and mitochondrial functionality.** **a**, Total protein content in indicated MEFs relative to WT/WT. **b**, Total RNA per cell for each of the indicated genotypes. **c**, **d**, WT/WT, G12D/WT and G12D/G12D MEFs were profiled by CASY counter (Roche) and cell volume (**c**) and diameter (**d**) measured. **a–d**, Mean value of three independent MEF triplicates per genotype  $\pm$  s.d. **e**, Oxygen consumption rate (OCR) of MEFs in response to oligomycin, CCCP and rotenone (two-way ANOVA). **f**, NAO staining was used to determine mitochondrial mass in *Kras*<sup>WT/WT</sup> (WT/WT), *Kras*<sup>G12D/WT</sup> (G12D/WT) and *Kras*<sup>G12D/G12D</sup> (G12D/G12D)

MEFs. Geometric mean of NAO fluorescence in cells was determined by FACS. Representative overlay (left panel) and geometric mean (right panel) displayed. **g**, Mitochondrial architecture was examined after Mitotracker green staining in WT/WT, G12D/WT and G12D/G12D MEFs (scale bar, 10  $\mu\text{m}$ ). **h**, TMRM staining was used to determine mitochondrial membrane potential in MEFs of indicated genotypes. Geometric mean of TMRM fluorescence in cells was determined by FACS. Representative overlay (left panel) and geometric mean (right panel) displayed. **e–h**, Representative data of three independent MEFs per genotype show mean of triplicates  $\pm$  s.d.; \*\*\* $P < 0.001$ ; one-way ANOVA.

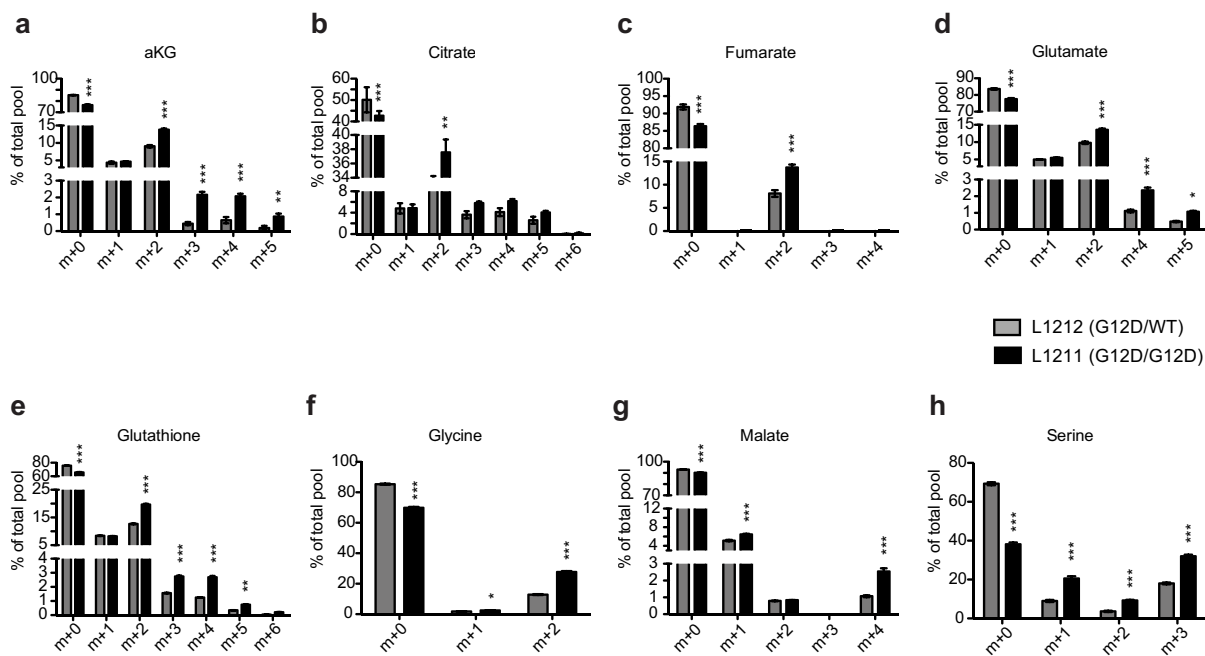
$^{13}\text{C}$ -GLC

**Extended Data Figure 3 | Glucose metabolism reprogramming in *Kras*<sup>G12D/G12D</sup> MEFs. a–j,** Measurement of  $^{13}\text{C}$ -glucose-derived metabolites, calculated as a percentage of the total metabolite pool following LC-MS analysis of WT/WT, G12D/WT and G12D/G12D MEFs

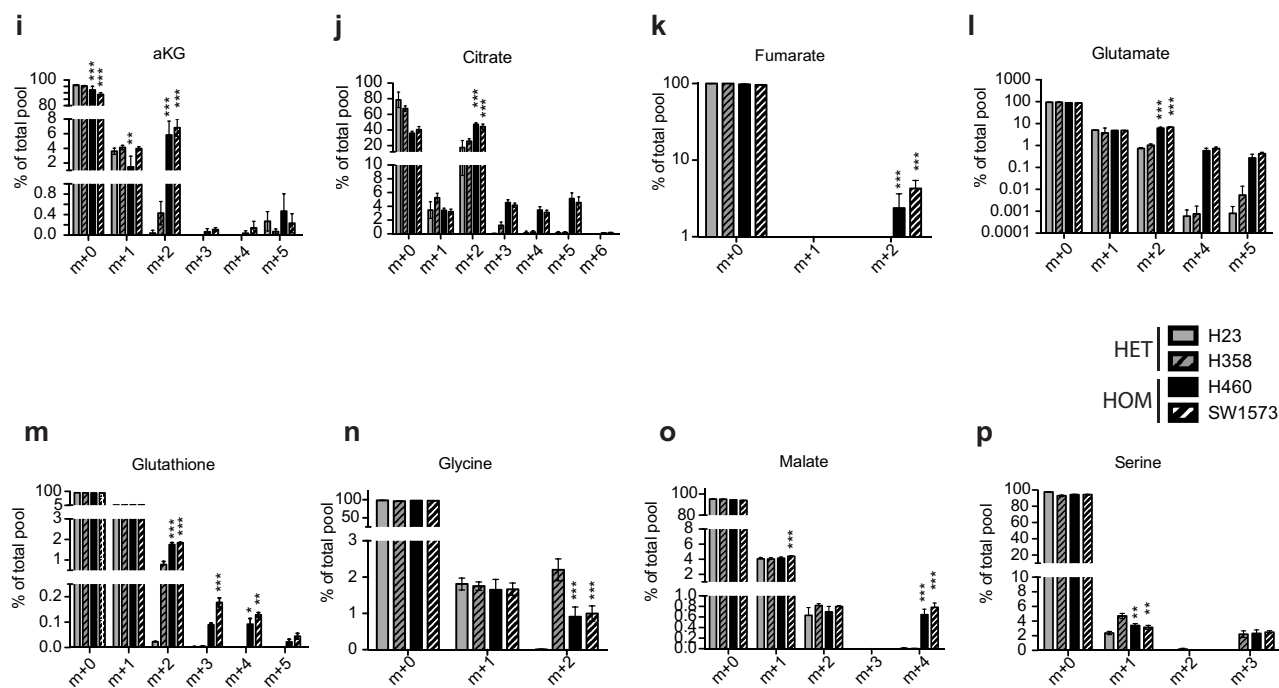
after 4 h culture with  $^{13}\text{C}$ -glucose-supplemented media. Representative data (of two independent MEFs per genotype) showing mean of triplicates  $\pm$  s.d.; \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$  (two-way ANOVA). Undetected isotopologues not shown.

$^{13}\text{C}$ -GLC

## Murine lung cancer cells



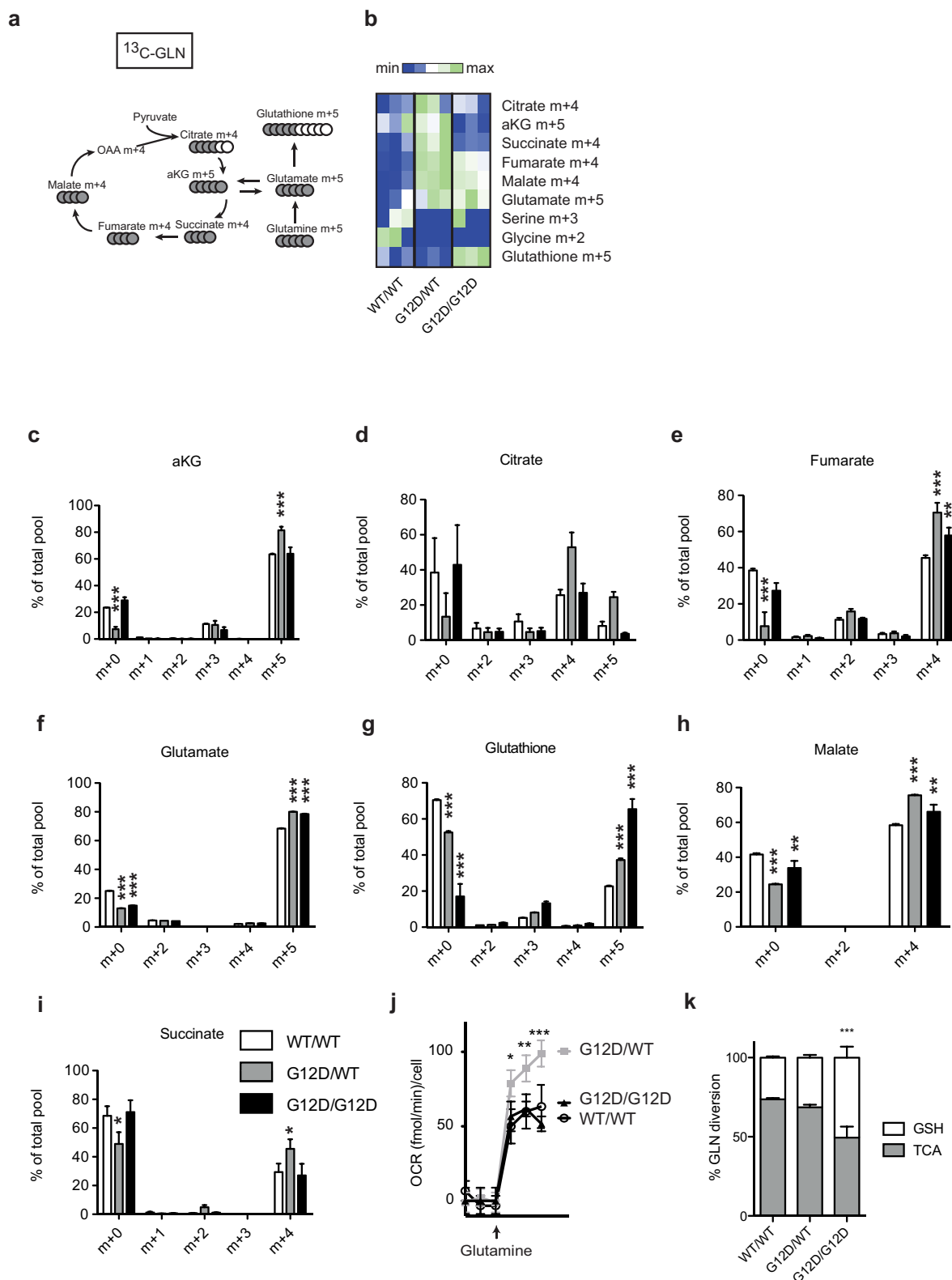
## Human lung cancer cells



**Extended Data Figure 4 | Glucose metabolism reprogramming in lung tumour cells with mutant *Kras* copy gain.** Measurement of  $^{13}\text{C}$ -glucose-derived metabolites, calculated as a percentage of the total metabolite pool following LC-MS analysis of murine (L1211 and L1212, a–h) and human (H23, H358, H460, SW1573, i–p) mutant *Kras* heterozygous and homozygous lung tumour cells. Cells were cultured

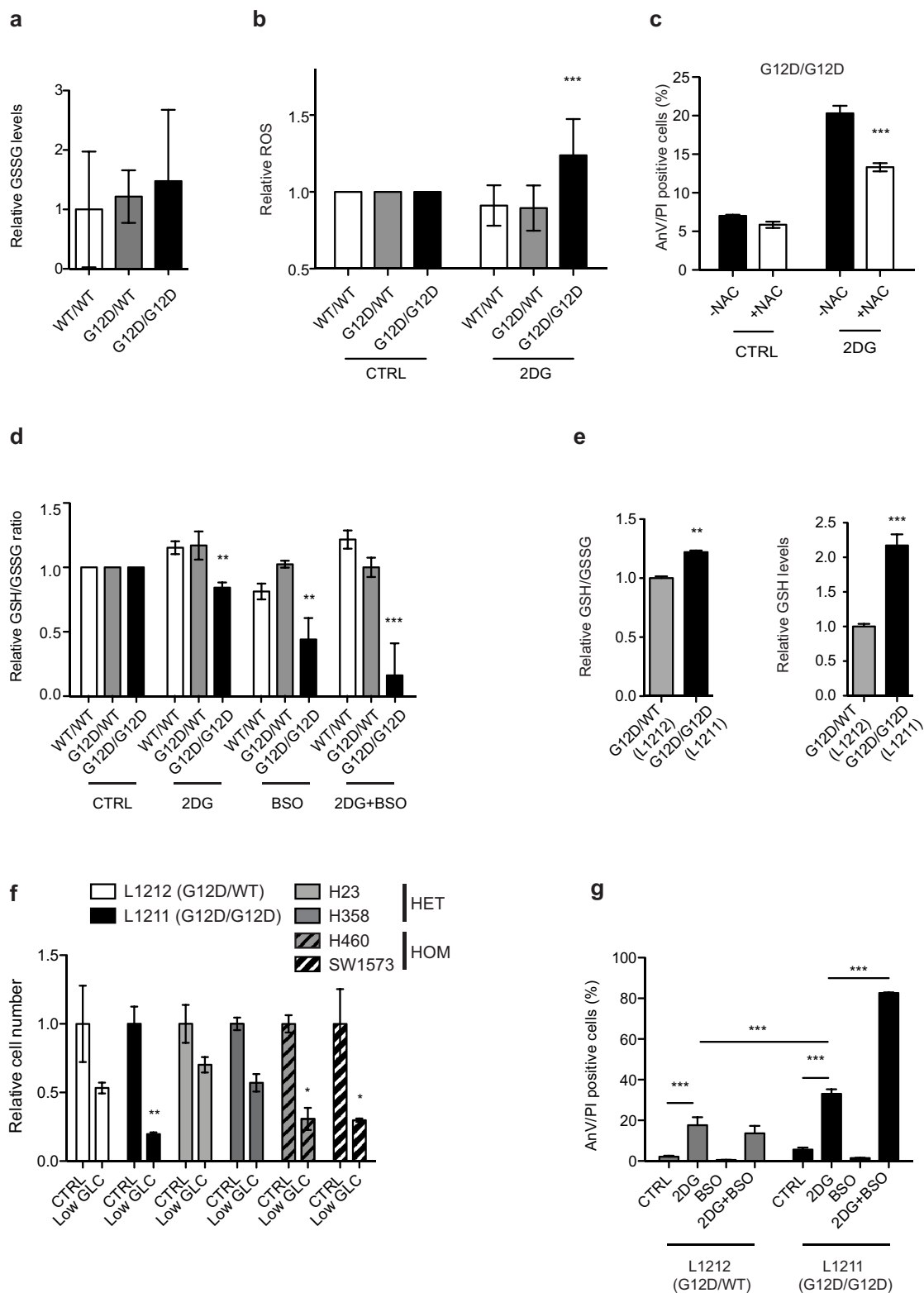
for 4 h with  $^{13}\text{C}$ -glucose-supplemented media before analysis. Data show mean of triplicates  $\pm$  s.d.; \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$  (two-way ANOVA, relative to *Kras*<sup>mut</sup> heterozygous cells; i–p, homozygous samples significantly different from both heterozygous cell lines indicated). Undetected isotopologues not shown.





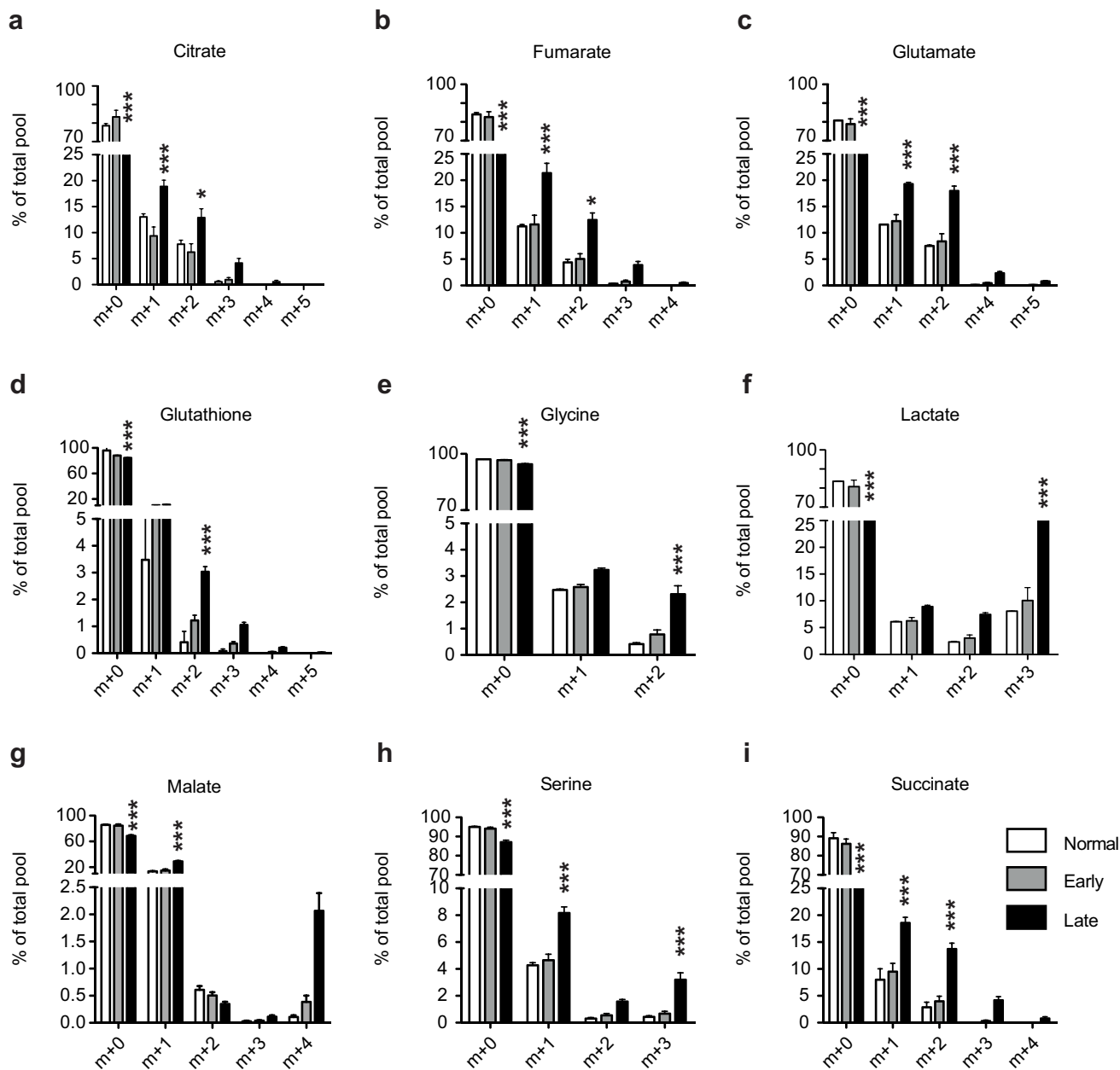
**Extended Data Figure 5 | *Kras*<sup>G12D/WT</sup> and *Kras*<sup>G12D/G12D</sup> MEFs have distinct glutamine metabolism profiles.** Glutamine metabolism analysis in WT/WT, G12D/WT and G12D/G12D MEFs. **a**, Representation of carbon flux (grey circles) from uniformly labelled  $^{13}\text{C}$ -glutamine ( $^{13}\text{C}$ -GLN). **b**, Heatmap illustrates abundance of selected labelled metabolites across triplicates of representative MEFs (two independent MEFs per genotype analysed) based on metabolomics analysis. **c–i**, Measurement of  $^{13}\text{C}$ -glutamine-derived metabolites, calculated as a percentage of the total metabolite pool following LC–MS analysis of WT/WT, G12D/WT and G12D/G12D MEFs after 4 h culture with  $^{13}\text{C}$ -glutamine-supplemented

media. Representative data (two independent MEFs per genotype) show mean of triplicates  $\pm$  s.d. (two-way ANOVA). **j**, Oxygen consumption rate (OCR) of WT/WT, G12D/WT and G12D/G12D MEFs upon glutamine (4 mM) addition. Representative data of three independent MEFs per genotype showing mean of triplicates  $\pm$  s.d. **k**, Relative diversion (percentage) of glutamine to TCA (aKG m + 5) or GSH (GSH m + 5) in MEFs of indicated genotypes based on metabolomics data. Representative MEF data ( $n = 2$  MEFs per genotype) show triplicate mean  $\pm$  s.d. (one-way ANOVA). \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$ .



**Extended Data Figure 6 | *Kras*<sup>G12D</sup> homozygous cells depend on glucose metabolism reprogramming for ROS management.** **a**, GSSG levels in G12D/WT and G12D/G12D MEFs relative to WT/WT. Mean data ( $n = 3$  MEFs per genotype)  $\pm$  s.d. shown. **b**, ROS levels in MEFs following 48 h of 2DG treatment. Data were normalized to vehicle treatment (CTRL). **c**, Percentage of AnnexinV/PI double-positive G12D/G12D MEFs following 48 h of 2DG treatment in the presence (+) or absence (-) of NAC. **d**, Ratio of reduced to oxidized glutathione (GSH/GSSG) determined for WT/WT, G12D/WT and G12D/G12D MEFs after incubation with 2DG, BSO or both (2DG + BSO) for 48 h, normalized to vehicle. **b–d**, Representative data from three independent MEFs per

genotype presented. Mean data for triplicates  $\pm$  s.d. shown (two-way ANOVA). **e**, Representative data of GSH/GSSG ratio and GSH levels in murine G12D/G12D tumour cells relative to G12D/WT ( $t$ -test). **f**, Differential sensitivity of lung tumour cells to nutrient depletion. Lung tumour cells were cultured in normal media and low glucose conditions for 72 h and viable cells counted and normalized to CTRL (two-way ANOVA). **g**, Percentage of AnnexinV/PI double-positive murine tumour cells following 48 h treatment with BSO, 2DG, or both (2DG + BSO). **e–g**, Representative data ( $n = 3$  independent experiments) depict triplicate mean  $\pm$  s.d. (\*\*\* $P < 0.001$ , two-way ANOVA). \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$ .

Lung tumours *in vivo* $^{13}\text{C}$ -GLC**Extended Data Figure 7 | Increased mutant *Kras* allelic content leads to glucose metabolism reprogramming in lung tumours *in vivo*.**

**a–i.** Control (no Cre) and tumour-bearing *Kras*<sup>G12D/+</sup>;*p53*<sup>Fx/Fx</sup> mice were infused with  $^{13}\text{C}$ -glucose 12 (early group) or 16 weeks (late group) after adenoviral-Cre treatment and individual lung tumours (early,  $n = 16$ ;

late,  $n = 12$ ) or control lung (normal,  $n = 3$ ) collected for LC-MS analysis (three technical replicates per sample). Selected  $^{13}\text{C}$ -glucose-derived metabolites shown, calculated as a percentage of the total metabolite pool. Mean abundance per cohort  $\pm$  s.e.m. shown. \*\*\* $P < 0.001$ ; \* $P < 0.05$  (two-way ANOVA).



**Extended Data Table 1 | Mutant heterozygous and homozygous *KRAS* NSCLC cell lines**

<i>KRAS</i> Heterozygous		<i>KRAS</i> Homozygous	
Cell line	<i>KRAS</i> Mutation	Cell line	<i>KRAS</i> Mutation
NCI-H23	p.G12C	NCI-H460	p.Q61H
NCI-H358	p.G12C	SW1573	p.G12C

*KRAS* mutation and zygosity of four NSCLC cell lines according to COSMIC Cell Lines project database (version 73). Genotypes were confirmed by Sanger sequencing (data not shown).

Extended Data Table 2 | *KRAS* status in panel of human lung adenocarcinomas

<i>KRAS</i> <sup>mut</sup>				<i>KRAS</i> <sup>mut&amp;CG</sup>			
ID	<i>KRAS</i> mutation	<i>KRAS</i> GISTIC	<i>KRAS</i> CNV	ID	<i>KRAS</i> mutation	<i>KRAS</i> GISTIC	<i>KRAS</i> CNV
TCGA-78-7539-01	G12C	0	-0.065	TCGA-95-7567-01	G12V	1	0.135
TCGA-78-7160-01	G12D	0	-0.032	TCGA-78-7148-01	G12C	1	0.255
TCGA-49-6761-01	G12D	0	-0.032	TCGA-55-7726-01	G12C	1	0.34
TCGA-50-5933-01	G12C	0	-0.027	TCGA-55-7911-01	G12V,K88*	1	0.358
TCGA-05-4403-01	G12C	0	-0.008	TCGA-05-4415-01	G12C	1	0.38
TCGA-35-3615-01	G12C	0	-0.008	TCGA-75-5126-01	G12C	1	0.399
TCGA-55-6970-01	G12V	0	-0.008	TCGA-05-4390-01	G12V	1	0.4
TCGA-75-7030-01	G12D	0	-0.005	TCGA-05-4418-01	G12C	1	0.404
TCGA-55-6983-01	G12C	0	0	TCGA-73-4662-01	G12C	1	0.411
TCGA-55-7725-01	G12V	0	0	TCGA-49-4510-01	G12D	1	0.413
TCGA-44-6146-01	G12V	0	0	TCGA-78-7161-01	G12V	1	0.481
TCGA-78-7540-01	G12V	0	0.001	TCGA-78-7145-01	G12Y	1	0.523
TCGA-67-3773-01	G12V	0	0.002	TCGA-86-7713-01	G12C	1	0.547
TCGA-97-7941-01	G12A	0	0.004	TCGA-05-4417-01	G12C	1	0.585
TCGA-49-4505-01	G12C	0	0.004	TCGA-91-6828-01	G12V	1	0.632
TCGA-78-7166-01	G12C	0	0.004	TCGA-69-7980-01	G12A	1	0.647
TCGA-44-7659-01	G12C	0	0.004	TCGA-55-7907-01	G12C	1	0.746
TCGA-44-6776-01	G12D	0	0.004	TCGA-55-7728-01	G12V	1	0.797
TCGA-50-7109-01	G12C	0	0.005	TCGA-75-7027-01	G12V	1	0.812
TCGA-50-5932-01	G12C	0	0.011	TCGA-50-5941-01	G12A	1	1.458
TCGA-55-7281-01	G12C	0	0.012	TCGA-64-5775-01	Q61L	1	1.752
TCGA-80-5608-01	G12A	0	0.019	TCGA-05-4395-01	G12V	1	2.25
TCGA-05-4249-01	G12C	0	0.024	TCGA-55-7815-01	G12V	2	0.491
TCGA-53-7813-01	G12A	0	0.032	TCGA-55-7283-01	G12S	2	0.64
TCGA-64-5774-01	G12C	0	0.034	TCGA-97-7554-01	G12V	2	0.836
TCGA-93-7347-01	G12C	0	0.036	TCGA-55-7576-01	G12S	2	0.875
TCGA-38-4626-01	G12V	0	0.044	TCGA-05-4433-01	G12V	2	1.082
TCGA-44-6777-01	G12C	0	0.045	TCGA-67-3774-01	G12F	2	1.095
TCGA-64-1677-01	G12C	0	0.051	TCGA-44-7672-01	G12A	2	2.604
TCGA-05-4430-01	G12C	0	0.056				
TCGA-49-6744-01	G12C	0	0.058				
TCGA-73-4659-01	G12V	0	0.069				
TCGA-50-5936-01	G12C	0	0.088				
TCGA-78-7167-01	G12F	0	0.093				
TCGA-55-6642-01	G12V	0	0.099				
TCGA-44-6145-01	G12V	0	0.099				

*KRAS* mutation, GISTIC score and copy number variation (CNV) in TCGA lung adenocarcinoma data set<sup>1</sup>. *KRAS* mutation (status and nucleotide substitution), *KRAS* putative copy number calls from GISTIC 2.0 analysis (GISTIC score) and *KRAS* copy number variation (Affymetrix SNP6) were downloaded from cBioportal. Mutant *KRAS* tumours with a GISTIC score of 0<sup>+</sup> ( $n = 65$ ) were divided into two cohorts: *KRAS* mutant (*KRAS*<sup>mut</sup>, GISTIC = 0,  $n = 36$ ) and *KRAS* mutant and copy gain (*KRAS*<sup>mut&CG</sup>, GISTIC = 1 & 2,  $n = 29$ ), as displayed.

# Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer

Alexandra C. Walls<sup>1\*</sup>, M. Alejandra Tortorici<sup>2,3\*</sup>, Berend-Jan Bosch<sup>4\*</sup>, Brandon Frenz<sup>1</sup>, Peter J. M. Rottier<sup>4</sup>, Frank DiMaio<sup>1</sup>, Félix A. Rey<sup>2,3</sup> & David Veasley<sup>1</sup>

**The tremendous pandemic potential of coronaviruses was demonstrated twice in the past few decades by two global outbreaks of deadly pneumonia. Entry of coronaviruses into cells is mediated by the transmembrane spike glycoprotein S, which forms a trimer carrying receptor-binding and membrane fusion functions<sup>1</sup>. S also contains the principal antigenic determinants and is the target of neutralizing antibodies. Here we present the structure of a mouse coronavirus S trimer ectodomain determined at 4.0 Å resolution by single particle cryo-electron microscopy. It reveals the metastable pre-fusion architecture of S and highlights key interactions stabilizing it. The structure shares a common core with paramyxovirus F proteins<sup>2,3</sup>, implicating mechanistic similarities and an evolutionary connection between these viral fusion proteins. The accessibility of the highly conserved fusion peptide at the periphery of the trimer indicates potential vaccinology strategies to elicit broadly neutralizing antibodies against coronaviruses. Finally, comparison with crystal structures of human coronavirus S domains allows rationalization of the molecular basis for species specificity based on the use of spatially contiguous but distinct domains.**

Coronaviruses are enveloped viruses responsible for 30% of mild respiratory infections and atypical pneumonia in humans worldwide<sup>4</sup>. The emergence of the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002 and of the Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 demonstrated that these zoonotic viruses can transmit to humans from various animal species, and suggested that additional emergence events are likely to occur. The fatality rate of SARS-CoV and MERS-CoV infections are about 10–37%<sup>1,4</sup> and there are no approved antiviral treatments or vaccines.

Coronaviruses use S homotrimers to promote cell attachment and fusion of the viral and host membranes. S determines host range, cell tropism and is the main target of neutralizing antibodies during infection<sup>1</sup>. S is a class I viral fusion protein synthesized as a single chain precursor of about 1,300 amino acids that trimerizes upon folding. It is composed of an amino-terminal S<sub>1</sub> subunit, containing the receptor-binding domain, and a carboxy-terminal S<sub>2</sub> subunit, driving membrane fusion. Cleavage by furin-like host proteases at the junction between S<sub>1</sub> and S<sub>2</sub> (S<sub>2</sub> cleavage site) occurs during biogenesis for some coronaviruses such as mouse hepatitis virus (MHV, the prototypical and best-studied coronavirus)<sup>1,5</sup>. The S<sub>1</sub> and S<sub>2</sub> subunits remain non-covalently associated in the metastable pre-fusion S trimer. After virion uptake by target cells, a second cleavage is mediated by endo-lysosomal proteases (S<sub>2</sub>' cleavage site), allowing fusion activation of coronavirus S proteins<sup>6</sup>.

Crystal structures of coronavirus S post-fusion cores demonstrated that the fusogenic conformational changes lead to the formation of a so-called trimer of hairpins that is the hallmark of class I fusion proteins<sup>7–10</sup>. These structures contain two heptad-repeat (HR) regions

present in S<sub>2</sub> assembled as an extended triple helical coiled-coil motif (HR1) surrounded by three shorter helices (HR2). Crystal structures of several coronavirus S receptor-binding domains in complex with their cognate receptors have also been reported<sup>11–14</sup>. Finally, cryo-electron microscopy (cryoEM) of SARS-CoV virions provided a snapshot of the S glycoprotein at 16 Å resolution<sup>15</sup>. The lack of high-resolution data for any coronavirus S trimer has prevented a detailed analysis of the infection mechanisms.

We produced an MHV S ectodomain trimer with enhanced stability by mutating the S<sub>2</sub> cleavage site and fusing a GCN4 trimerization motif at the C-terminal end of the construct. The resulting MHV S ectodomain forms a trimer binding with high-affinity to the soluble mouse CEACAM1a receptor (Extended Data Fig. 1a, b). We used state-of-the-art cryoEM<sup>16</sup> to determine the structure of the MHV S ectodomain trimer at 4.0 Å resolution (Fig. 1a–c and Extended Data Figs 2 and 3). We fitted the crystal structures of two S<sub>1</sub> domains<sup>11,13,17</sup> and built *de novo* the rest of the polypeptide chain using Coot<sup>18</sup> and Rosetta<sup>19,20</sup> (Fig. 1d–f, Extended Data Figs 2–4 and Supplementary Tables 1 and 2). The final model includes residues 15 to 1118, with an internal break corresponding to a loop immediately upstream from the S<sub>2</sub>' cleavage site (residues 827–863). The region connecting the S<sub>1</sub> and S<sub>2</sub> subunits (residues 718–754) features weak density that correlates with its accessibility for proteolytic cleavage *in vivo*. Residues 453–535 were modelled by density-guided homology modelling using Rosetta owing to the poor quality of the density in this region (Extended Data Fig. 3k).

The MHV S ectodomain is a 140 Å long trimer with a triangular cross-section varying in diameter from 70 Å, at the membrane proximal base, to 140 Å at the membrane distal end (Fig. 1d, e). The structure comprises two functional subunits (Fig. 2a–d): a distal moiety constituted by the S<sub>1</sub> subunits; and a central stem connecting to the viral membrane formed by the S<sub>2</sub> subunits.

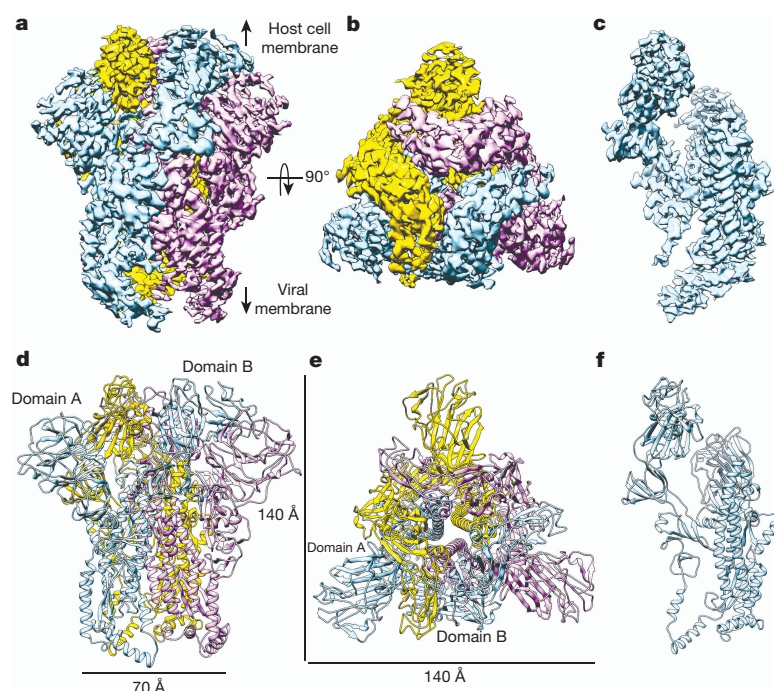
The S<sub>1</sub> subunit has a 'V' shape contributing to the overall triangular appearance of the S trimer (Extended Data Fig. 5a). The S<sub>1</sub> N-terminal moiety comprises domain A, which is folded as a galectin-like β-sandwich decorated with extended loops on the viral membrane distal side, and a three-stranded antiparallel β-sheet plus an α-helix on the viral membrane proximal side. The S<sub>1</sub> C-terminal half folds as three spatially distinct β-rich domains, termed B, C and D (Fig. 2a–d).

The S<sub>2</sub> subunit connects to the viral membrane and is characterized by the presence of long α-helices (Figs 2b–d and 3a). A central helix (α<sub>30</sub>) stretches 75 Å along the three-fold molecular axis towards the viral membrane (Fig. 3a). It is located immediately downstream of the HR1 motif, which folds as four consecutive α-helices (α<sub>26</sub>–α<sub>29</sub>; Fig. 3a and Extended Data Fig. 6a, b), in sharp contrast to the 120-Å long HR1 helix observed in the post-fusion S structures<sup>7–9</sup> (Extended Data Fig. 6c–e). The 55-Å-long upstream helix (α<sub>20</sub>), so named because it is located immediately upstream of the S<sub>2</sub>' cleavage site, runs parallel to and is zipped against the central helix via hydrophobic contacts

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Institut Pasteur, Unité de Virologie Structurale, 75015 Paris, France. <sup>3</sup>CNRS UMR 3569 Virologie, 75015 Paris, France. <sup>4</sup>Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, The Netherlands.

\*These authors contributed equally to this work.





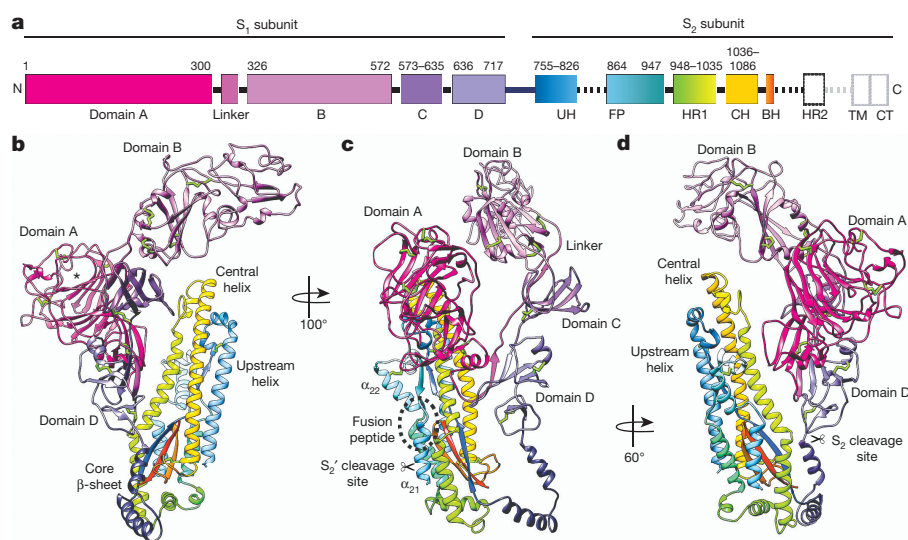
**Figure 1 | 3D reconstruction of the MHV S trimer determined by single-particle cryoEM.** **a–c**, 3D map filtered at 4.0 Å resolution coloured by protomer. Two different views of the S trimer (from the side (**a**) and from the top, looking towards the viral membrane (**b**)), and a side view of one S protomer (**c**) are shown. **d–f**, Ribbon diagrams showing the MHV S atomic model oriented as in **a–c**.

largely following a heptad-repeat pattern. A core antiparallel  $\beta$ -sheet ( $\beta_{46}$ – $\beta_{49}$ – $\beta_{50}$ ) is present at the viral membrane proximal end and is assembled from an N-terminal  $\beta$ -strand ( $\beta_{46}$ ), preceding the upstream helix, and a C-terminal  $\beta$ -hairpin ( $\beta_{49}$ – $\beta_{50}$ ), located downstream of the central helix.

MHV  $S_2$  features a topology similar to the paramyxovirus F proteins (such as respiratory syncytial virus (RSV) F: root mean squared deviation (r.m.s.d.) 4 Å over 125 residues), with a comparable 3D organization of the core  $\beta$ -sheet, the upstream helix and the central helix (Fig. 3a, b). Importantly, these motifs were shown to remain invariant in the pre- and post-fusion F structures<sup>2,3</sup>. The conservation of these motifs among coronavirus S and paramyxovirus F proteins suggests that these fusion proteins have evolved from a distant common ancestor. Although the density is too weak to trace the polypeptide chain downstream from  $\beta_{50}$ , secondary structure predictions suggest that the domain directly preceding HR2 could adopt a similar fold in coronavirus S and paramyxovirus F proteins.

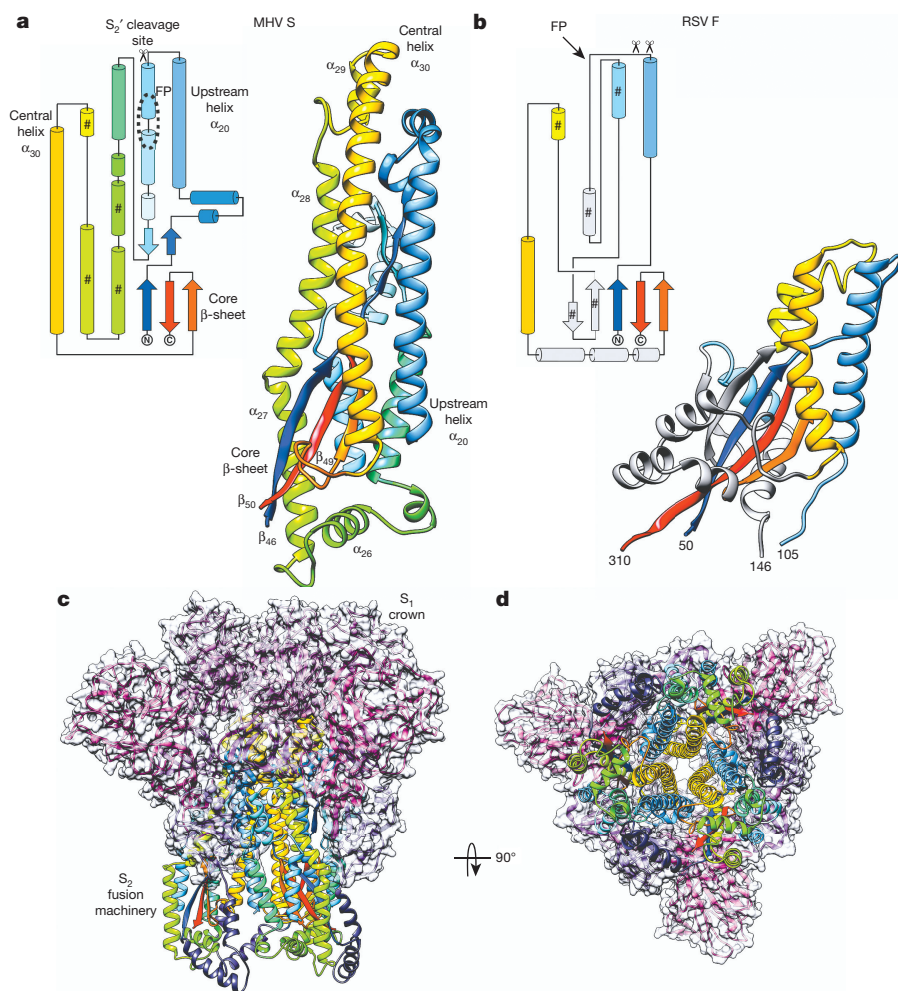
In the S trimer, the three central helices are packed via their central portions whereas the two ends splay away from the three-fold axis (Extended Data Fig. 7a–c). Additional contacts between the

upstream and central helices participate to inter-protomer interactions. Furthermore, the  $S_1$  subunits interlock to form a crown around the  $S_2$  trimer stabilizing it in the pre-fusion conformation (Fig. 3c, d and Supplementary Table 3). This is illustrated by the large surface area buried at the interface between each  $S_1$  subunit and the  $S_2$  subunits of the three protomers (1,970 Å<sup>2</sup>). Many of these contacts involve the HR1 helices and the fusion peptide region. These polypeptide segments undergo major refolding during the fusogenic conformational changes (Extended Data Fig. 6a–e), which supports the notion that the  $S_1$  subunits maintain the  $S_2$  fusion machinery in its metastable state. Substitutions of the conserved alanine 994 by valine in helix  $\alpha_{28}$  or of the conserved leucine 1062 by phenylalanine in the central helix were shown to attenuate fusogenicity<sup>21,22</sup>. Our structure suggests that the former substitution would strengthen hydrophobic packing against the core  $\beta$ -sheet (Extended Data Fig. 7b), and that the later substitution could reinforce molecular stapling of the central helices (Extended Data Fig. 7a, c). The expected modification of the energy landscape between pre-fusion and post-fusion conformations would explain the reduction in fusion activity of these mutants<sup>21,22</sup>.



**Figure 2 | Architecture of the MHV S protomer.**

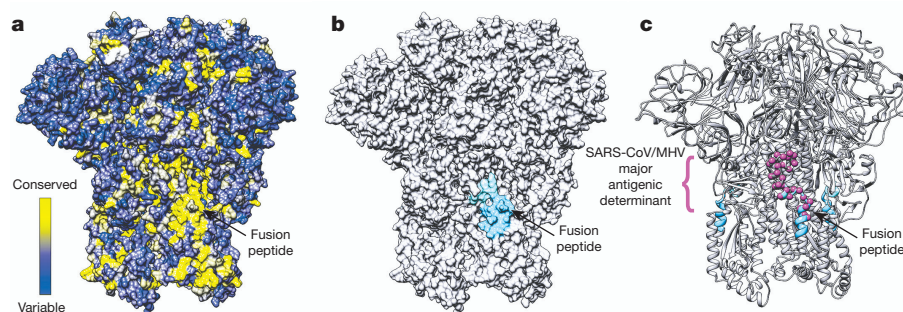
**a**, Schematic diagram of the S glycoprotein organization. Black and grey dashed lines denote regions unresolved in the reconstruction and regions that were not part of the construct, respectively. BH,  $\beta$ -hairpin ( $\beta_{49}$ – $\beta_{50}$ ); CH, central helix; CT, cytoplasmic tail; FP, fusion peptide; HR1/HR2, heptad-repeats; TM, transmembrane domain; UH, upstream helix. **b–d**, Ribbon diagrams depicting three views of the S protomer coloured as in **a**. Asterisk denotes the MHV S receptor-binding region. Disulfide bonds are shown as green sticks except for residues 453–535, for which they are not shown.



**Figure 3 | Pre-fusion structure of the coronavirus fusion machinery.** **a, b**, Topology and ribbon diagrams showing the structural similarity between coronavirus MHV S<sub>2</sub> (starting at residue 755) (**a**) and paramyxovirus RSV F (PDB 5C6B) (**b**). For clarity, only part of RSV F is shown, with conserved secondary structural elements coloured identically as for MHV S<sub>2</sub>. ‘#’ denotes motifs participating to the post-fusion HR1 coiled-coil. **c, d**, Two different views of the MHV S trimer (from the side (**c**) and top, looking towards the host cell membrane (**d**)) highlighting how S<sub>1</sub> (ribbons and semi-transparent surface) wraps around the S<sub>2</sub> fusion machinery (ribbons) to stabilize it.

The predicted fusion peptide includes the C-terminal half of helix  $\alpha_{21}$  and extends up to the N-terminal half of  $\alpha_{22}$  (refs 6 and 23) (Fig. 2c).  $\alpha_{21}$  is an amphipathic helix located at the periphery of the S trimer, burying hydrophobic side chains towards the S<sub>2</sub> centre and exposing charged residues to solvent (Fig. 2c and Extended Data Fig. 7b, c). In the case of porcine epidemic diarrhoea coronavirus, trypsin processing at the S<sub>2</sub>' site can only occur after host cell attachment<sup>24</sup>. This indicates that receptor binding could allosterically increase the accessibility of the S<sub>2</sub>' site, which is located within helix  $\alpha_{21}$ . The acidic pH of the endolysosomes could also contribute to exposing the S<sub>2</sub>' cleavage site for coronaviruses requiring cleavage in this compartment. The fact that helix  $\alpha_{21}$  appears dynamic and is found immediately downstream from

a disordered loop suggests that it could undergo considerable ‘breathing’ motions. Regardless of the mechanism promoting cleavage, the MHV S structure reported here explains the requirement for processing at the S<sub>2</sub>' site, as it frees the fusion peptide from the S<sub>2</sub> N-terminal region, which is a prerequisite for its insertion ~200 Å away in the target membrane. The peripheral position of the fusion peptide is similar to what has been observed in the parainfluenza virus 5 F<sup>3</sup> and HIV gp41 (ref. 25) prefusion structures (Extended Data Fig. 8a–c). The notable accessibility of the fusion peptide and its sequence conservation among coronaviruses<sup>6,23</sup> suggest that it would be an ideal target for epitope-focused vaccinology initiatives aimed at raising broadly neutralizing antibodies against S glycoproteins (Fig. 4a–c and Extended Data Fig. 9). Major



**Figure 4 | Potential strategy for neutralizing coronavirus infections.** **a**, Surface representation of the MHV S trimer coloured according to sequence conservation using the alignment presented in Extended Data Fig. 9. The fusion peptide sequence is highly conserved among coronavirus S proteins. **b**, Surface representation of the MHV S trimer

highlighting the peripheral position of the fusion peptide (blue and cyan). **c**, Ribbon diagrams of the MHV S trimer showing the overlapping positions of the fusion peptide (residues 870–887, blue and cyan) and of a major antigenic determinant identified for MHV and SARS-CoV (residues 875–905, magenta spheres).



antigenic determinants (inducing neutralizing antibodies) of MHV and SARS-CoV S proteins overlap with the fusion peptide region and support the suitability of this approach<sup>26,27</sup>. Antibodies binding to this site will not only hinder insertion of the fusion peptide into the target membrane, but will also putatively prevent fusogenic conformational changes. This epitope-focused strategy has proven successful to obtain neutralizing antibodies against RSV F<sup>28</sup>.

The spatial proximity of domains A and B in the S trimer allows rationalization of their alternative use among coronaviruses to interact with host receptors. MHV uses the viral membrane distal loops decorating domain A to interact with CEACAM1a (ref. 13), whereas MERS-CoV and SARS-CoV rely on the  $\beta$ -motif protruding from domain B to bind to DPP4 (ref. 11) or ACE2 (refs 12 and 14), respectively (Extended Data Fig. 5a–d). The poor sequence conservation of the B domain  $\beta$ -motif among coronavirus S proteins, its considerable length variation among MHV strains (Extended Data Fig. 9) and our density-guided homology model of this motif indicate structural and functional differences. These structural variations constitute the molecular basis underlying coronavirus species specificity and cell tropism using a single S architectural scaffold.

Sequence comparisons indicate that the MHV spike S<sub>1</sub> and S<sub>2</sub> subunits respectively share ~25% and ~40% sequence similarity with many other coronavirus S proteins (Extended Data Fig. 9). Therefore, the structure reported here is representative of the architecture of other coronavirus S such as those of MERS-CoV and SARS-CoV. This hypothesis is further supported by the structural similarity of (1) the MHV<sup>13</sup> and bovine coronavirus<sup>17</sup> A domains; (2) the MHV, MERS-CoV<sup>11</sup>, SARS-CoV<sup>12</sup> and HKU4 (ref. 29) B domains (Extended Data Fig. 10); (3) the post-fusion cores of MHV<sup>7</sup>, SARS-CoV<sup>8,10</sup> and MERS-CoV<sup>9</sup>; and (4) the isolation of infectious coronaviruses featuring a deletion of the A domain and using domain B as the receptor-binding domain<sup>30</sup>. Our results now provide a framework to understand coronavirus entry and suggest ways for preventing or treating future coronavirus outbreaks.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 17 November 2015; accepted 13 January 2016.**

**Published online 8 February 2016.**

- Du, L. *et al.* The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nature Rev. Microbiol.* **7**, 226–236 (2009).
- McLellan, J. S. *et al.* Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113–1117 (2013).
- Yin, H. S., Wen, X., Paterson, R. G., Lamb, R. A. & Jardetzky, T. S. Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. *Nature* **439**, 38–44 (2006).
- Coleman, C. M. & Frieman, M. B. Coronaviruses: important emerging human pathogens. *J. Virol.* **88**, 5209–5212 (2014).
- Bosch, B. J., van der Zee, R., de Haan, C. A. & Rottier, P. J. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **77**, 8801–8811 (2003).
- Burkard, C. *et al.* Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. *PLoS Pathog.* **10**, e1004502 (2014).
- Xu, Y. *et al.* Structural basis for coronavirus-mediated membrane fusion. Crystal structure of mouse hepatitis virus spike protein fusion core. *J. Biol. Chem.* **279**, 30514–30522 (2004).
- Duquerry, S., Vigouroux, A., Rottier, P. J., Rey, F. A. & Bosch, B. J. Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein. *Virology* **335**, 276–285 (2005).
- Gao, J. *et al.* Structure of the fusion core and inhibition of fusion by a heptad repeat peptide derived from the S protein of Middle East respiratory syndrome coronavirus. *J. Virol.* **87**, 13134–13140 (2013).
- Supekar, V. M. *et al.* Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein. *Proc. Natl Acad. Sci. USA* **101**, 17958–17963 (2004).
- Lu, G. *et al.* Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* **500**, 227–231 (2013).

- Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
- Peng, G. *et al.* Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc. Natl Acad. Sci. USA* **108**, 10696–10701 (2011).
- Wu, K., Li, W., Peng, G. & Li, F. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proc. Natl Acad. Sci. USA* **106**, 19970–19974 (2009).
- Beniac, D. R., Andonov, A., Grudski, E. & Booth, T. F. Architecture of the SARS coronavirus prefusion spike. *Nature Struct. Mol. Biol.* **13**, 751–752 (2006).
- Cheng, Y. Single-particle cryo-EM at crystallographic resolution. *Cell* **161**, 450–457 (2015).
- Peng, G. *et al.* Crystal structure of bovine coronavirus spike protein lectin domain. *J. Biol. Chem.* **287**, 41931–41938 (2012).
- Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).
- DiMaio, F. *et al.* Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nature Methods* **12**, 361–365 (2015).
- Wang, R. Y. *et al.* De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nature Methods* **12**, 335–338 (2015).
- Krueger, D. K., Kelly, S. M., Lewicki, D. N., Ruffolo, R. & Gallagher, T. M. Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *J. Virol.* **75**, 2792–2802 (2001).
- Taguchi, F. & Matsuyama, S. Soluble receptor potentiates receptor-independent infection by murine coronavirus. *J. Virol.* **76**, 950–958 (2002).
- Madu, I. G., Roth, S. L., Belouzard, S. & Whittaker, G. R. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J. Virol.* **83**, 7411–7421 (2009).
- Wicht, O. *et al.* Proteolytic activation of the porcine epidemic diarrhea coronavirus spike fusion protein by trypsin in cell culture. *J. Virol.* **88**, 7952–7961 (2014).
- Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455–461 (2014).
- Daniel, C. *et al.* Identification of an immunodominant linear neutralization domain on the S2 portion of the murine coronavirus spike glycoprotein and evidence that it forms part of complex tridimensional structure. *J. Virol.* **67**, 1185–1194 (1993).
- Zhang, H. *et al.* Identification of an antigenic determinant on the S2 domain of the severe acute respiratory syndrome coronavirus spike glycoprotein capable of inducing neutralizing antibodies. *J. Virol.* **78**, 6938–6945 (2004).
- Correia, B. E. *et al.* Proof of principle for epitope-focused vaccine design. *Nature* **507**, 201–206 (2014).
- Wang, Q. *et al.* Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* **16**, 328–337 (2014).
- Reguera, J. *et al.* Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. *PLoS Pathog.* **8**, e1002859 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** Research reported in this publication was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under Award Number T32GM008268 (A.C.W.). Part of this research was facilitated by the National Resource for Automated Molecular Microscopy (award number GM103310), and the Hyak supercomputer system at the University of Washington. We thank W. Bartelink, B. Tummers and B. van der Kooij for assistance in cloning, protein expression and characterization. We are also grateful to B. Baron for assistance with the Thermophoresis experiments.

**Author Contributions** M.A.T., B.-J.B., P.J.M.R., F.A.R. and D.V. designed the experiments. B.-J.B. and P.J.M.R. designed and cloned the protein constructs. B.-J.B. and M.A.T. carried out protein expression, purification and biophysical characterization. D.V. performed cryoEM sample preparation and data collection. A.C.W. and D.V. processed the cryoEM data. A.C.W., B.F., F.D. and D.V. built the atomic model. A.C.W., M.A.T., B.F., B.-J.B., F.D., F.A.R. and D.V. analysed the data. A.C.W., F.A.R. and D.V. prepared the manuscript, with input from all authors.

**Author Information** The cryoEM map and the atomic model have been deposited in the Electron Microscopy Data Bank (EMD) and the Protein Data Bank (PDB) under accession codes EMD-6526 and 3JCL, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.V. (dveesler@uw.edu) or F.A.R. (felix.rey@pasteur.fr).



## METHODS

No statistical methods were used to predetermine sample size.

**Plasmids.** A human codon-optimized gene encoding the MHV spike gene (UniProt: P11224) was synthesized with an Arg717Ser amino acid mutation to abolish the furin cleavage site at the S<sub>1</sub>-S<sub>2</sub> junction (S<sub>2</sub> cleavage site). From this gene, the fragment encoding the MHV ectodomain (residues 15–1231) was PCR-amplified and ligated to a gene fragment encoding a GCN4 trimerization motif (IKRMKQIEDKIEIESKQKKIENEIARIKKIK)<sup>3,31</sup>, a thrombin cleavage site (LVPRGSLE), an 8-residue long Strep-Tag (WSHPQFEK) and a stop codon. This construct results in fusing the GCN4 trimerization motif in register with the HR2 helix at the C-terminal end of the MHV S-encoding sequence. This gene was cloned into the pMT/BiP/V5/His expression vector (Invitrogen) in frame with the *Drosophila* BiP secretion signal downstream the metallothionein promoter. The D1 domain of mouse CEACAM1a (residues 35–142; gb NP\_001034274.1) was amplified by PCR and cloned into a mammalian expression plasmid<sup>32</sup>, in frame with a CD5 signal sequence at the 5' end, and with a sequence encoding a thrombin cleavage site, a glycine linker and the Fc domain of human IgG1 at the 3' end, creating the pCD5-MHVR-T-Fc vector.

**Production of recombinant CEACAM1a ectodomain by transient transfection.**

293-F cells were grown in suspension using FreeStyle 293 Expression Medium (Life technologies) at 37 °C in a humidified 5% CO<sub>2</sub> incubator on a Celltron shaker platform (Infors HT) rotating at 130 r.p.m. (for 1 l culture flasks). Twenty-four hours before transfection, cell density was adjusted at  $1.5 \times 10^6$  cells ml<sup>-1</sup>, and culture grown overnight in the same conditions as mentioned above to reach  $\sim 2.5 \times 10^6$  cells ml<sup>-1</sup> the day of transfection. Cells were collected by centrifugation at 1,250 r.p.m. for 5 min, and resuspended in fresh FreeStyle 293 Expression Medium (Life technologies) without antibiotics at a density of  $2.5 \times 10^6$  cells ml<sup>-1</sup>.

To produce recombinant CEACAM1a ectodomain, 400 µg of pCD5-MHVR-T-Fc vector (purified using EndoFree plasmid kit from Qiagen) were added to 200 ml of suspension cells. The cultures were swirled for 5 min on shaker in the culture incubator before adding 9 µg ml<sup>-1</sup> of Linear polyethylenimine (PEI) solution (25 kDa, Polysciences). Twenty-four hours after transfection, cells were diluted 1:1 with FreeStyle 293 Expression Medium and the transfected cells were cultivated for 6 days. Clarified cell supernatants were concentrated tenfold using Vivaflow tangential filtration cassettes (Sartorius, 10-kDa cut-off) before affinity purification using a Protein A column (GE LifeSciences) followed by gel filtration chromatography using a Superdex 200 10/300 GL column (GE Life Sciences) equilibrated in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl. The Fc tag was removed by trypsin cleavage in a reaction mixture containing 7 mg of recombinant CEACAM1a ectodomain and 5 µg of trypsin in 100 mM Tris-HCl, pH 8.0 and 20 mM CaCl<sub>2</sub>. The reaction mixture was incubated at 25 °C overnight and re-loaded in a Protein A column to remove uncleaved protein and the Fc tag. The cleaved protein was further purified by gel filtration using a Superdex 75 column 10/300 GL (GE Life Sciences) equilibrated in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl. The purified protein was quantified using absorption at 280 nm and concentrated to approximately 10 mg ml<sup>-1</sup>.

**Production of recombinant MHV S ectodomain in *Drosophila* S2 cells.** To generate a stable *Drosophila* S2 cell line expressing recombinant MHV spike ectodomain, we used Effectene (Qiagen) and 2 µg of the plasmid encoding the MHV S protein ectodomain. A second plasmid, encoding blasticidin S deaminase was cotransfected as dominant selectable marker. Stable MHV S ectodomain expressing cell lines were selected by addition of 10 µg ml<sup>-1</sup> blasticidin S (Invivogen) to the culture medium 48 h after transfection.

For large-scale production of MHV S ectodomain the cells were cultured in spinner flasks and induced by 5 µM CdCl<sub>2</sub> at a density of approximately 10<sup>7</sup> cells per ml. After a week at 28 °C, clarified cell supernatants were concentrated 40-fold using Vivaflow tangential filtration cassettes (Sartorius, 10-kDa cut-off) and adjusted to pH 8.0, before affinity purification using StrepTactin Superflow column (IBA) followed by gel filtration chromatography using Superose 6 10/300 GL column (GE Life Sciences) equilibrated in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl. The purified protein was quantified using absorption at 280 nm and concentrated to approximately 4 mg ml<sup>-1</sup>.

**SEC-MALS.** For size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) analysis, samples (0.2 ml at 1 mg ml<sup>-1</sup>) were loaded onto a Superdex 200 10/300 GL column (GE Life Sciences, 0.4 ml min<sup>-1</sup> in gel filtration buffer) and passed through a Wyatt DAWN Heleos II EOS 18-angle laser photometer coupled to a Wyatt Optilab TrEX differential refractive index detector. Data were analysed using Astra 6 software (Wyatt Technology Corp).

**MicroScale Thermophoresis.** Solution MicroScale Thermophoresis (MST) binding studies were performed using standard protocols on a Monolith NT.115 (Nanotemper Technologies). In brief, recombinant CEACAM1a ectodomain protein was labelled using the RED-NHS (Amine Reactive) Protein Labelling Kit (Nanotemper Technologies). The MHV S ectodomain protein was serially

diluted in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl and the labelled recombinant CEACAM1a was added to a final concentration of 500 nM before overnight incubation at 4 °C. The CEACAM1a concentration was chosen such that the observed fluorescence was approximately 1,000 U at 40% LED power. The samples were loaded into standard-treated Monolith capillaries and were measured by standard protocols using a Monolith NT.115, NanoTemper. The changes in the fluorescent thermophoresis signal were plotted against the concentration of the serially diluted MHV spike protein, and K<sub>d</sub> values were determined using the NanoTemper analysis software.

**CryoEM sample preparation and data collection.** Three microlitres of MHV spike at 1.85 mg ml<sup>-1</sup> was applied to a 1.2/1.3 C-flat grid (Protochips), which had been glow-discharged for 30 s at 20 mA. Thereafter, grids were plunge-frozen in liquid ethane using a Gatan CP3 and a blotting time of 3.5 s. Data were acquired using an FEI Titan Krios transmission electron microscope operated at 300 kV and equipped with a Gatan K2 Summit direct detector. Coma-free alignment was performed using the Leginon software<sup>33</sup>. Automated data collection was carried out using Leginon<sup>34</sup> to control both the FEI Titan Krios (used in microprobe mode at a nominal magnification of 22,500×) and the Gatan K2 Summit operated in counted mode (pixel size: 1.315 Å) at a dose rate of  $\sim 9$  counts per physical pixel per s, which corresponds to  $\sim 12$  electrons per physical pixels per s (when accounting for coincidence loss<sup>35</sup>). Each video had a total accumulated exposure of 53 e<sup>-</sup> Å<sup>-2</sup> fractionated in 38 frames of 200 ms (yielding movies of 7.6 s). A data set of  $\sim 1,600$  micrographs was acquired in a single session using a defocus range of between 2.0 and 5.0 µm.

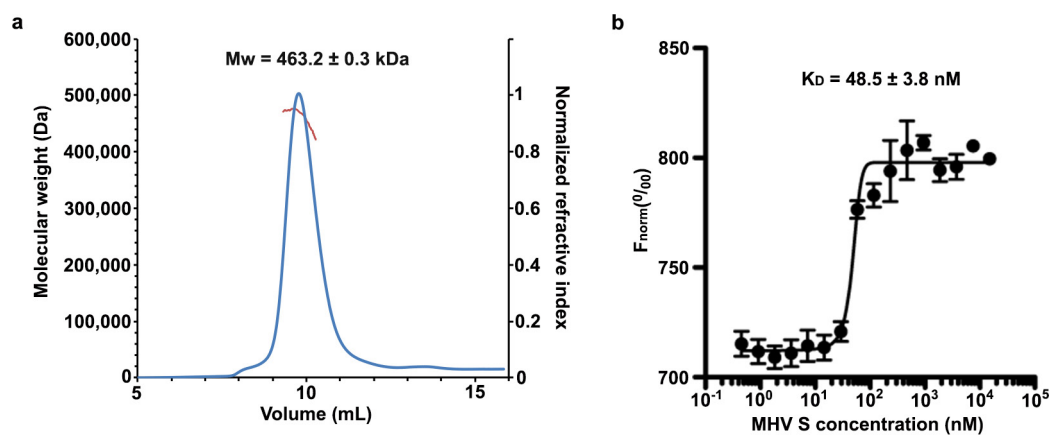
**CryoEM data processing.** Whole-frame alignment was carried out using the software developed previously<sup>35</sup>, which is integrated into the Appion pipeline<sup>36</sup>, to account for stage drift and beam-induced motion. The parameters of the micro-scope contrast transfer function were estimated for each micrograph using ctffind3 (ref. 37). Micrographs were manually masked using Appion to exclude the visible carbon supporting film for further processing. Particles were automatically picked in a reference-free manner using DogPicker<sup>38</sup>. Extraction of particle images was performed using Relion 1.4 with a box size of 320 pixels<sup>2</sup> and applying a windowing operation in Fourier space to yield a final box size of 288 pixels<sup>2</sup> (corresponding to a pixel size of 1.46 Å). From the 1.2 million particles initially picked, a subset of 50,000 particles were randomly selected to generate class averages using RELION<sup>39</sup>. An initial 3D model was generated using OPTIMOD<sup>40</sup> within the Appion pipeline. The entire data set was subjected to 2D alignment and clustering using RELION and particles belonging to the best-defined class averages were retained ( $\sim 500,000$  particles). These  $\sim 500,000$  particles were then subjected to RELION 3D classification with four classes (using c1 symmetry) starting with our initial model low-pass filtered to 40 Å resolution. We subsequently used the  $\sim 230,000$  best particles (selected from the 3D classification) and the map corresponding to the best 3D class (low-pass filtered at 40 Å resolution) to run Relion 3D auto-refine (c3 symmetry), which led to a reconstruction at 4.4 Å resolution. We used the particle polishing procedure in RELION 1.4 to correct for individual particle movement and radiation damage<sup>41,42</sup>. A second round of 3D classification with 6 classes (c3 symmetry) was performed using the polished particles resulting in the selection of 82,000 particles. A new 3D auto-refine run (c3 symmetry) using the selected 82,000 particles and the map corresponding to the best 3D class (low-pass filtered at 40 Å resolution) yielded a map at 4.0 Å resolution following post-processing in RELION. The final map was sharpened with an empirically determined B factor of  $-220$  Å<sup>2</sup> using Relion post processing. Reported resolutions are based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion<sup>43</sup>, and Fourier shell correction curves were corrected for the effects of soft masking by high-resolution noise substitution<sup>44</sup>. The soft mask used for FSC calculation had a 10 pixel cosine edge fall-off. The overall shape and dimensions of our reconstruction agree with previous data although the HR2 stem connecting to the membrane is not resolved<sup>15</sup>.

**Model building and analysis.** Fitting of atomic models into cryoEM maps was performed using UCSF Chimera<sup>45</sup> and Coot<sup>18,46</sup>. We initially docked the MHV domain A structure (PDB 3R4D) and used a crystal structure of a bovine coronavirus domain A (PDB 4H14) to model the three-stranded β-sheet and the α-helix present on the viral membrane proximal side of the galectin-like domain. Next, the MERS-CoV domain B crystal structure (PDB 4KQZ) was also fit into the density, and rebuilt and refined using RosettaCM<sup>47</sup>. Although we could accurately align the sequences corresponding to the core β-sheet of the MHV and MERS-CoV B domains, the  $\sim 100$  residues forming the β-motif extension (residues 453–535, MERS-CoV/SARS-CoV receptor-binding moiety) could not be aligned with confidence. We used RosettaCM to build models of each of the 945 possible disulfide patterns into the density for domain B. For each disulfide arrangement, 50 models were generated, and there was a very clear energy signal for a single such arrangement (Extended Data Fig. 3k). Then, 1,000 models with this disulfide arrangement were sampled, and the lowest energy model (using the Rosetta force field

augmented with a fit-to-density score term) was selected. Owing to the poor quality of the reconstruction at the apex of the S trimer, the confidence of the model is lowest for the segment corresponding to residues 453–535, as homology modelling was used to fill in details missing in the map.

A backbone model was then manually built for the rest of the S polypeptide using Coot. Sequence register was assigned by visual inspection where side chain density was clearly visible. This initial hand built model was used as an initial model for Rosetta *de novo*<sup>20</sup>. The Rosetta-derived model largely agreed with the hand-built model. Rosetta *de novo* successfully identified fragments allowing to anchor the sequence register for domains C and D as well as for helices  $\alpha_{21}$ – $\alpha_{25}$ . Given these anchoring positions, RosettaCM<sup>47</sup> augmented with a novel density-guided model-growing protocol was able to rebuild domains C and D in full. The final model was refined by applying strict non-crystallographic symmetry constraints using Rosetta<sup>19</sup>. Model refinement was performed using a training map corresponding to one of the two maps generated by the gold-standard refinement procedure in Relion. The second map (testing map) was used only for calculation of the FSC compared to the atomic model and preventing overfitting<sup>48</sup>. The quality of the final model was analysed with Molprobity<sup>49</sup>. Structure analysis was assisted by the PISA<sup>50</sup> and DALI<sup>51</sup> servers. The sequence alignment was generated using MultAlin<sup>52</sup> and coloured with ESPript<sup>53</sup>. All figures were generated with UCSF Chimera<sup>45</sup>.

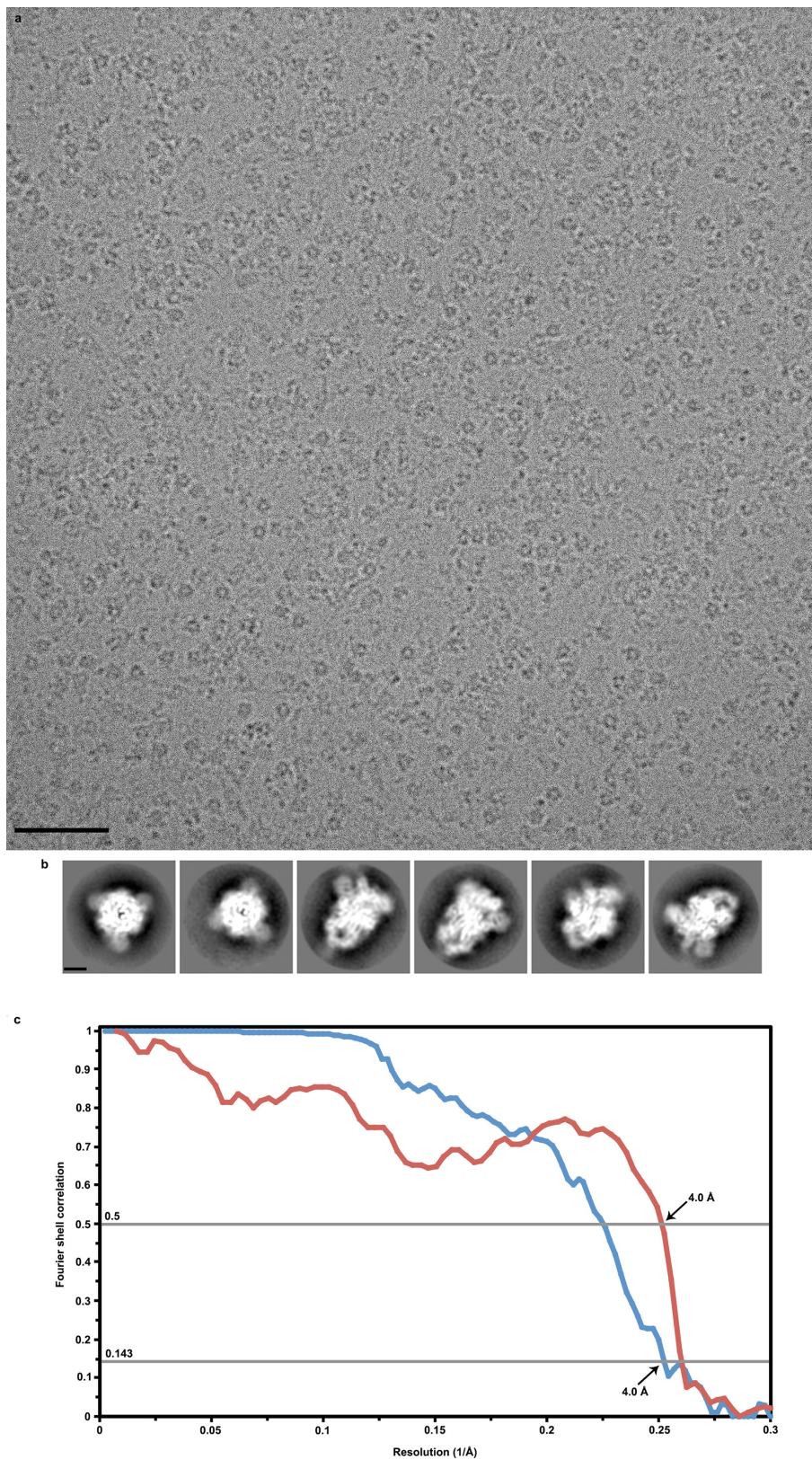
31. Eckert, D. M., Malashkevich, V. N. & Kim, P. S. Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. *J. Mol. Biol.* **284**, 859–865 (1998).
32. Zeng, Q., Langereis, M. A., van Vliet, A. L., Huizinga, E. G. & de Groot, R. J. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl Acad. Sci. USA* **105**, 9065–9069 (2008).
33. Glaeser, R. M., Typke, D., Tiemeijer, P. C., Pulokas, J. & Cheng, A. Precise beam-tilt alignment and collimation are required to minimize the phase error associated with coma in high-resolution cryo-EM. *J. Struct. Biol.* **174**, 1–10 (2011).
34. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
35. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
36. Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
37. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
38. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).
39. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
40. Lyumkis, D., Vinterbo, S., Potter, C. S. & Carragher, B. Optimod—an automated approach for constructing and optimizing initial models for single-particle electron microscopy. *J. Struct. Biol.* **184**, 417–426 (2013).
41. Bai, X. C., Fernandez, I. S., McMullan, G. & Scheres, S. H. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2**, e00461 (2013).
42. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).
43. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
44. Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
45. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
47. Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).
48. DiMaio, F., Zhang, J., Chiu, W. & Baker, D. Cryo-EM model validation using independent map reconstructions. *Protein Sci.* **22**, 865–868 (2013).
49. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
50. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
51. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
52. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890 (1988).
53. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).



**Extended Data Figure 1 | Biophysical characterization of the MHV S ectodomain.** **a**, The MHV S molecular mass was determined to be  $463.2 \pm 0.3$  kDa (mean  $\pm$  s.e.m.) (corresponding to a trimer) using size-exclusion chromatography coupled in-line with multi-angle light scattering and refractometry. The blue line represents the normalized refractive index (right ordinate axis) and the red line shows the estimated

molecular mass (expressed in Da, left ordinate axis). **b**, MHV S binds with high-affinity to the soluble mouse CEACAM1a receptor. Thermophoresis signal plotted against the MHV S concentration. The dissociation constant ( $K_d$ ) was determined to be  $48.5 \pm 3.8$  nM. Values correspond to the average of two independent experiments. The concentration of CEACAM1a used was 500 nM.

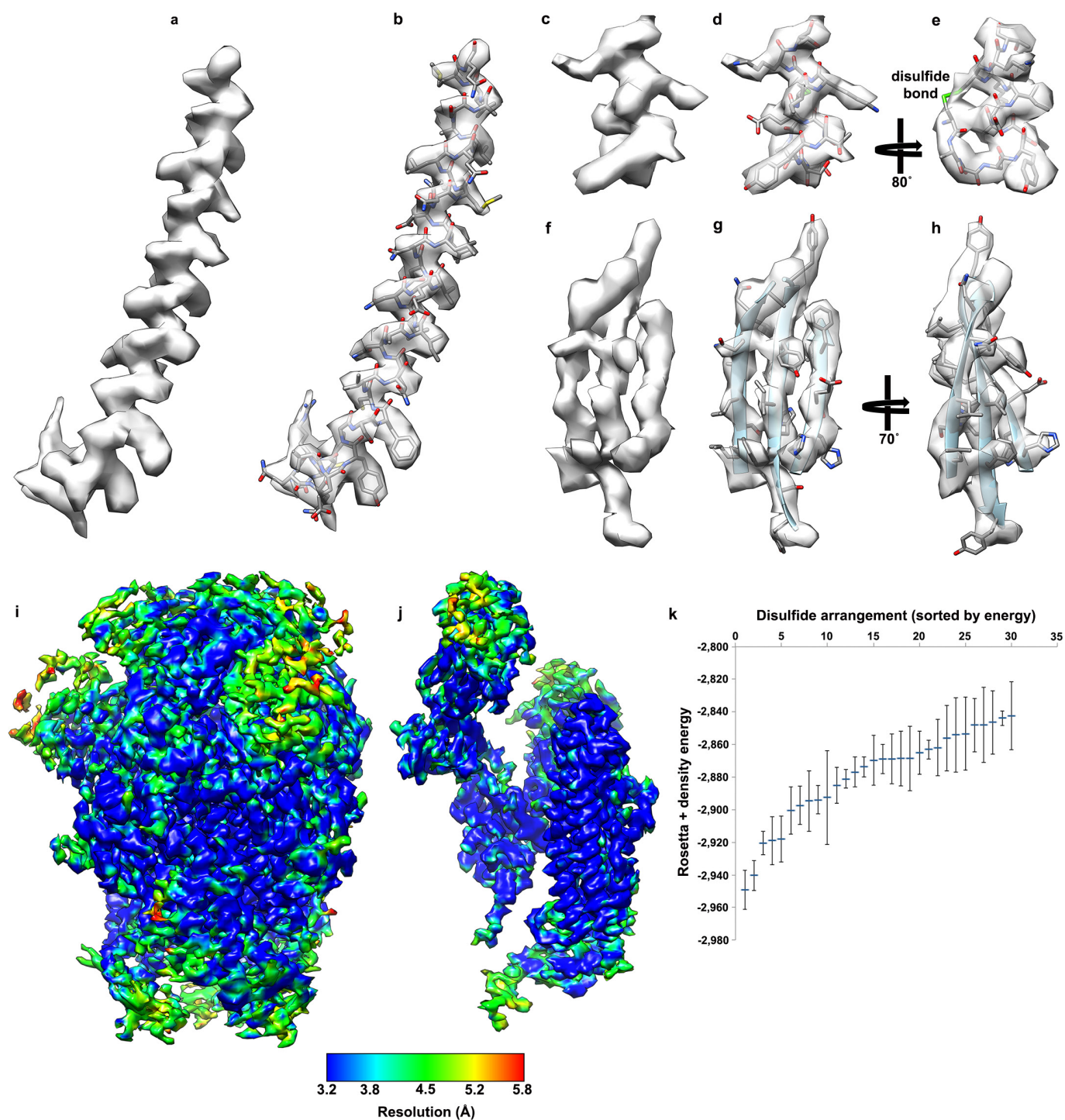




### Extended Data Figure 2 | CryoEM analysis of the MHV S trimer.

**a, b,** Representative electron micrograph (defocus: 4.6  $\mu\text{m}$ ) (**a**) and class averages (**b**) of the MHV S trimer embedded in vitreous ice. Scale bars: 573 Å (micrograph) and 44 Å (class averages). **c,** Gold-standard (blue) and

model/map (red) Fourier shell correlation (FSC) curves. The resolution was determined to be 4.0 Å. The 0.143 and 0.5 cut-off values are indicated by horizontal grey bars.



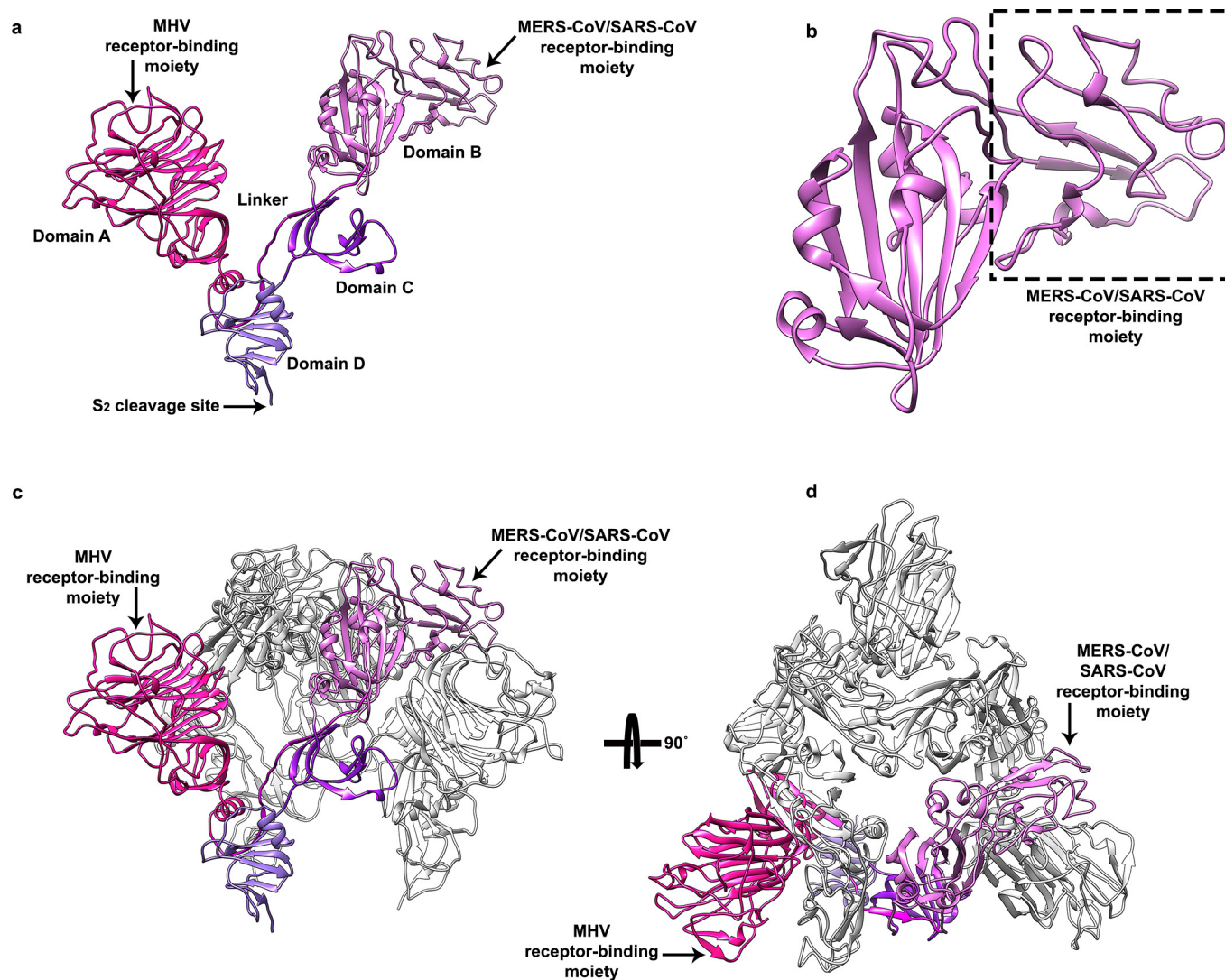
**Extended Data Figure 3 | CryoEM density for selected regions of the MHV S reconstruction, local resolution analysis and density-guided homology modelling of residues 453–535.** The atomic model is shown with the corresponding region of the map. **a, b**, Upstream helix. **c–e**, Helix belonging to domain A (residues 284–296). **f–h**, Core β-sheet. **i, j**, CryoEM density corresponding to the MHV S trimer (**i**) and a single

protomer (**j**), coloured according to local resolution determined with the software Resmap. We interpret Resmap results as a qualitative (rather than quantitative) estimate of map quality. **k**, Rebuilding of the MHV S domain B using RosettaCM. Plot showing the energy mean and s.d. of the models corresponding to the 30 lowest energy disulfide arrangements (out of 945) for domain B.

<b>Data collection</b>	
Number of particles	82,000
Pixel size (Å)	1.315 (rescaled to 1.46)
Defocus range (µm)	2-5
Voltage (kV)	300
Electron dose (e <sup>-</sup> /Å <sup>2</sup> )	53
<b>Refinement</b>	
Resolution	4.0
Map sharpening B factor (Å <sup>2</sup> )	-220
<b>Model validation</b>	
Molprobity score (percentile)	1.55 (94 <sup>th</sup> )
All-atom clashscore (percentile)	3.68 (97 <sup>th</sup> )
Poor rotamers (%)	0
Ramachandran favored (%)	94.26
Ramachandran allowed (%)	99.44
Ramachandran outliers (%)	0.56
r.m.s.d bonds (Å)	0.018
r.m.s.d angles (°)	1.8

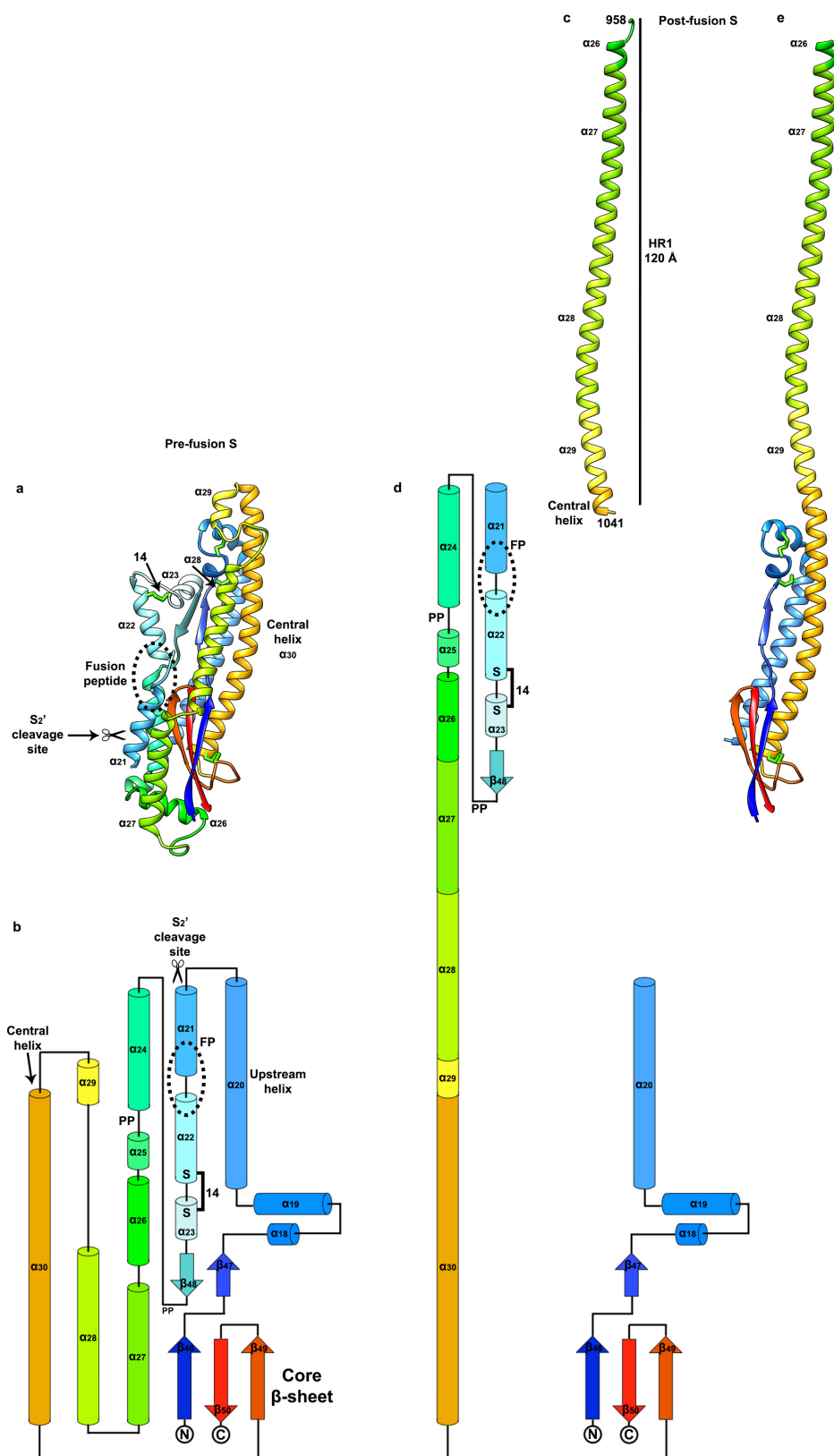
Extended Data Figure 4 | Refinement and model statistics.





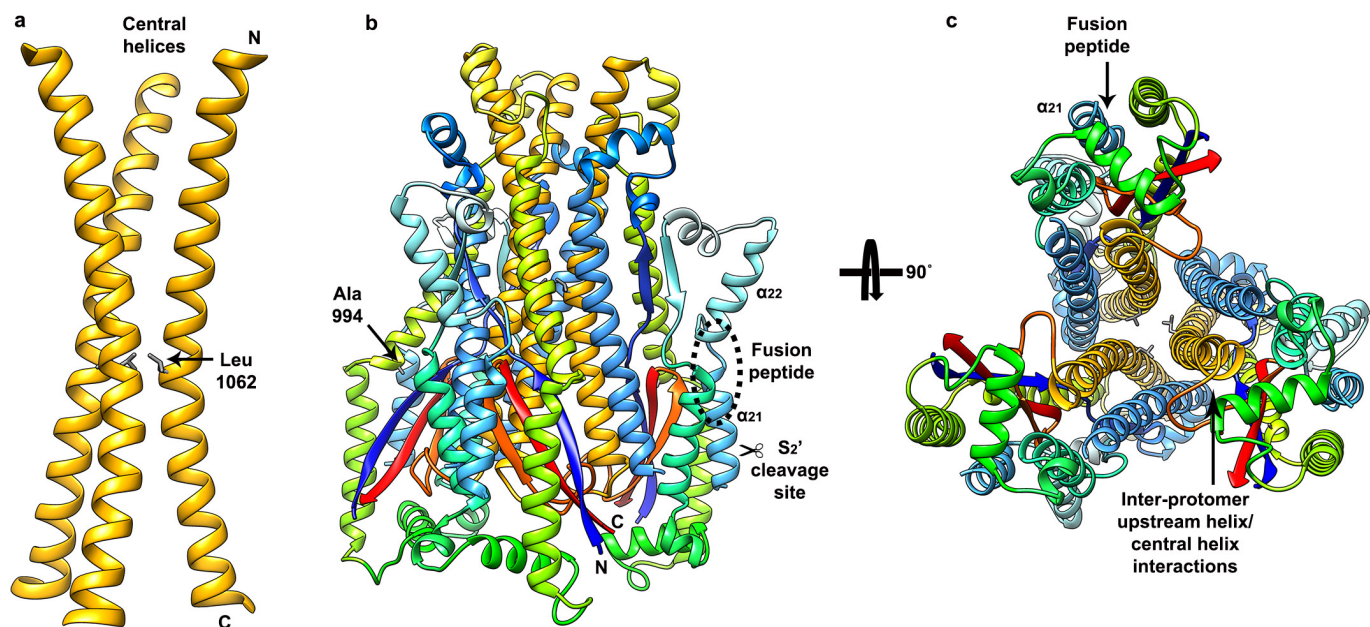
**Extended Data Figure 5 | Structural organization of the  $S_1$  subunit.** **a**, Ribbon diagram showing a single  $S_1$  protomer. **b**, Close-up view of the MHV S domain B. The structural motif used as a receptor-interacting moiety by MERS-CoV and SARS-CoV is indicated. The density was too

weak to allow tracing of this segment (residues 453–535), which has been traced by density-guided homology modelling using Rosetta. **c**, **d**, Ribbon diagrams of the  $S_1$  trimer viewed from the side (**c**) and from the top (looking towards the viral membrane) (**d**).



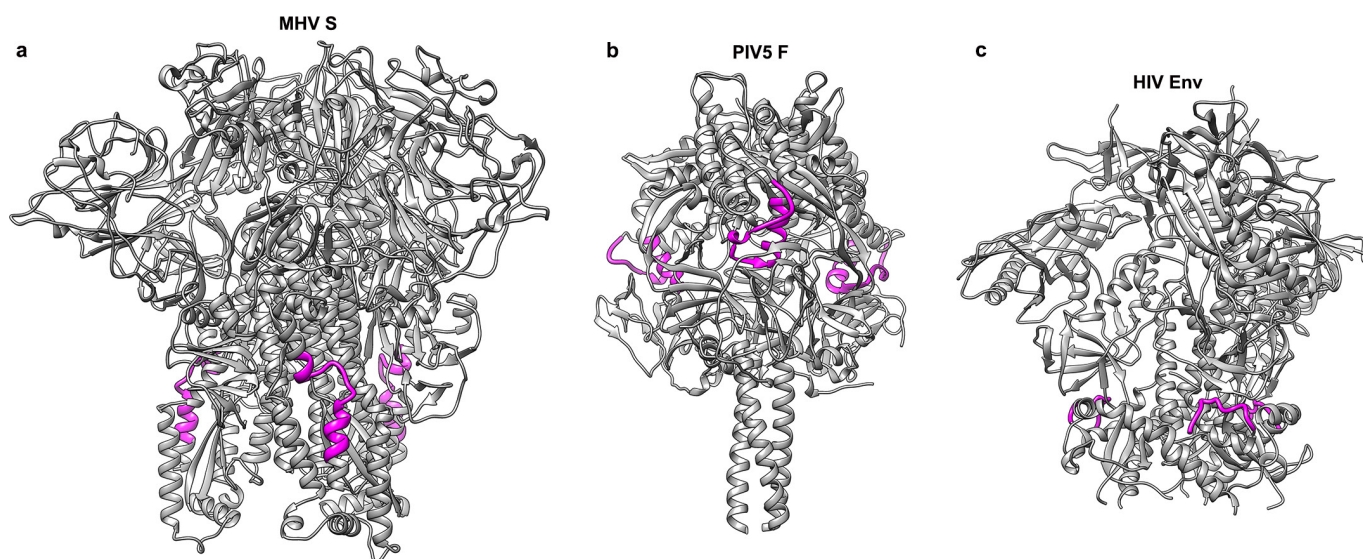
**Extended Data Figure 6 | Mechanisms of membrane fusion promoted by coronavirus S glycoproteins.** **a**, Ribbon diagram of the MHV S<sub>2</sub> pre-fusion structure. Disulfide bonds are shown as green sticks. **b**, Topology diagram of the MHV S<sub>2</sub> pre-fusion structure. PP, di-proline that will act as a helix breaker. The presence of these di-proline motifs indicates that the post-fusion HR1 coiled-coil could not extend up to the fusion peptide as a single helix. This hypothesis is further supported by the observation of a conserved disulfide bond formed between residues Cys894 and Cys905 (labelled 14 in **a** and **b**), which will prevent refolding of helices  $\alpha_{22}$  and

$\alpha_{23}$  as a single extended helix. **c**, Ribbon diagram of the SARS-CoV post-fusion HR1 helix obtained by X-ray crystallography (PDB 1WYY). The residue numbers corresponding to the MHV A59 sequence are indicated. **d**, Topology diagram showing the expected coronavirus S post-fusion conformation derived from our MHV S structure and the SARS-CoV post-fusion core crystal structure shown in **c**. **e**, Ribbon diagram of a model of the MHV S<sub>2</sub> post-fusion conformation. Residues belonging to  $\alpha_{21}$ ,  $\alpha_{22}$ ,  $\alpha_{23}$ ,  $\beta_{48}$ ,  $\alpha_{24}$  and  $\alpha_{25}$  are not represented owing to a lack of structural information.

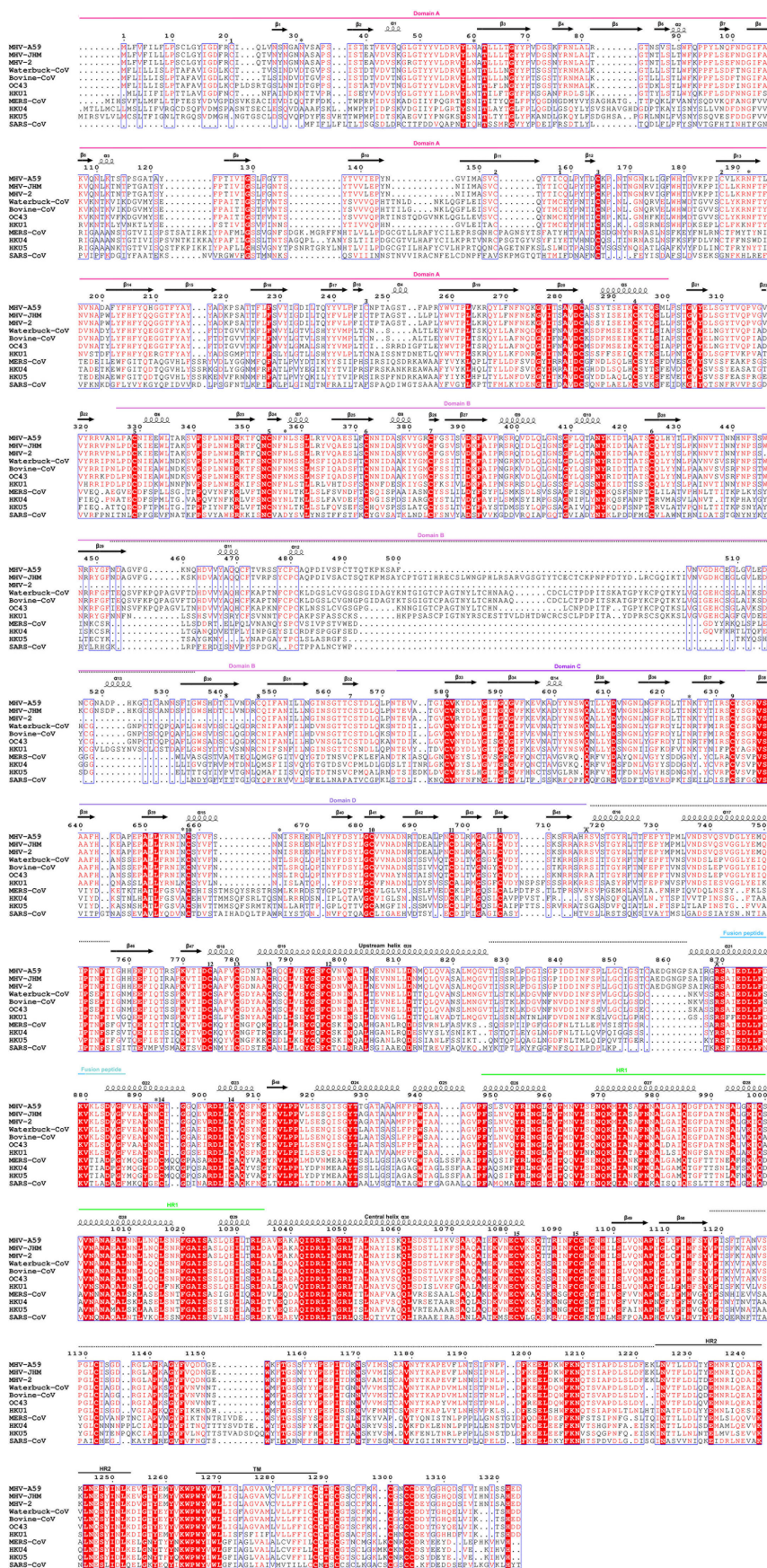


**Extended Data Figure 7 | Structural organization of the S<sub>2</sub> fusion machinery.** **a**, Ribbon diagram of the trimer of central helices. **b**, **c**, Ribbon diagrams of the S<sub>2</sub> trimer (starting at residue 755) viewed from the side (**b**) and from the bottom (looking towards the host cell membrane) (**c**). Residues Ala994 and Leu1062, which are discussed in the text, are shown in stick format.





**Extended Data Figure 8 | Class I viral fusion proteins with exposed fusion peptide.** **a**, MHV S (residues 870–887). **b**, Parainfluenza virus 5 F (PIV5 F, residues 103–128, PDB 2B9B). **c**, HIV-1 gp41 (residues 518–528, PDB 4TVP). The trimeric fusion proteins are shown as grey ribbon diagrams with the fusion peptides rendered in magenta.

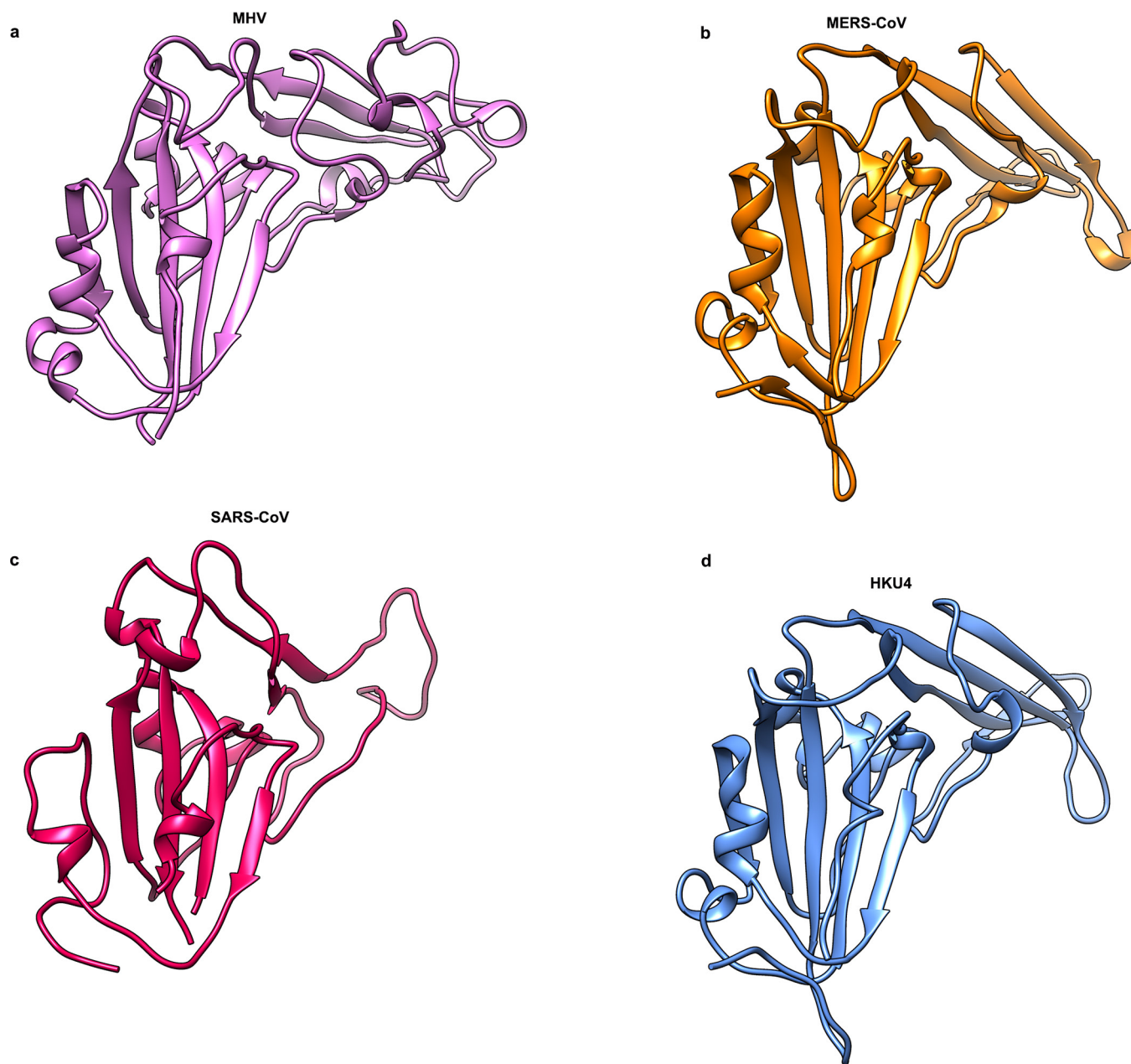


Extended Data Figure 9 | See next page for figure caption.

**Extended Data Figure 9 | Sequence conservation among coronavirus S glycoproteins. a,** Sequence alignment of coronavirus S proteins. Bovine-CoV, bovine respiratory coronavirus AH187 (gi 253756585); HKU1, human coronavirus HKU1 (gi 545299280); HKU4, tytonycteris bat coronavirus HKU4 (gi 126030114); HKU5, pipistrellus bat coronavirus HKU5 (gi 126030124); MERS-CoV, Middle East respiratory syndrome coronavirus (gi 836600681); MHV-A59, mouse hepatitis virus A59 (gi 1352862); MHV-JHM, mouse hepatitis virus JHM (gi 60115395); MHV-2, mouse hepatitis virus 2 (gi 5565844); OC43, human coronavirus OC43 (gi 744516696); SARS-CoV, severe acute respiratory syndrome coronavirus ZJ01 (gi 39980889); Waterbuck-CoV, waterbuck coronavirus

US/OH-WD358-TC/1994 (gi 215478096). Asparagine residues featuring N-linked glycan chains visible in the MHV S reconstruction are indicated with a star. The S<sub>2</sub> and S<sub>2</sub>' cleavage sites are indicated with scissors at positions corresponding to the MHV S sequence. Cysteine residues involved in the formation of disulfide bonds are numbered according to Supplementary Table 2. The secondary structure elements observed in our MHV S reconstruction are indicated above the sequence. The black dotted lines above the sequence indicate regions poorly defined in the density. Although the viral membrane distal loops of the A domains are weakly defined in the density, the availability of a crystal structure of this domain from the same virus (PDB 3R4D) helped with the modelling.





**Extended Data Figure 10 | Structural similarity of B domains among coronavirus S glycoproteins. a,** MHV (pink). **b,** MERS-CoV (orange, PDB 4KQZ). **c,** SARS-CoV (red, PDB 2AJF). **d,** HKU4 (blue, PDB 4QZV).

# Pre-fusion structure of a human coronavirus spike protein

Robert N. Kirchdoerfer<sup>1\*</sup>, Christopher A. Cottrell<sup>1\*</sup>, Nianshuang Wang<sup>2</sup>, Jesper Pallesen<sup>1</sup>, Hadi M. Yassine<sup>3†</sup>, Hannah L. Turner<sup>1</sup>, Kizzmekia S. Corbett<sup>3</sup>, Barney S. Graham<sup>3</sup>, Jason S. McLellan<sup>2</sup> & Andrew B. Ward<sup>1</sup>

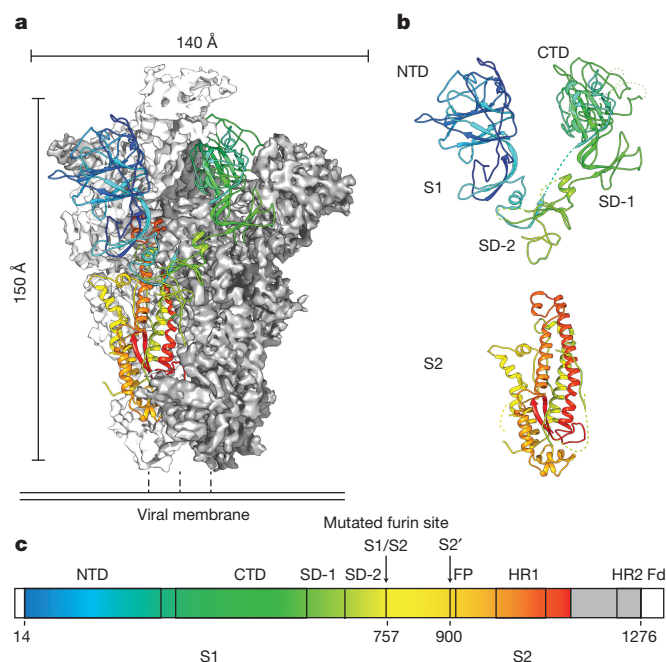
**HKU1 is a human betacoronavirus that causes mild yet prevalent respiratory disease<sup>1</sup>, and is related to the zoonotic SARS<sup>2</sup> and MERS<sup>3</sup> betacoronaviruses, which have high fatality rates and pandemic potential. Cell tropism and host range is determined in part by the coronavirus spike (S) protein<sup>4</sup>, which binds cellular receptors and mediates membrane fusion. As the largest known class I fusion protein, its size and extensive glycosylation have hindered structural studies of the full ectodomain, thus preventing a molecular understanding of its function and limiting development of effective interventions. Here we present the 4.0 Å resolution structure of the trimeric HKU1 S protein determined using single-particle cryo-electron microscopy. In the pre-fusion conformation, the receptor-binding subunits, S1, rest above the fusion-mediating subunits, S2, preventing their conformational rearrangement. Surprisingly, the S1 C-terminal domains are interdigitated and form extensive quaternary interactions that occlude surfaces known in other coronaviruses to bind protein receptors. These features, along with the location of the two protease sites known to be important for coronavirus entry, provide a structural basis to support a model of membrane fusion mediated by progressive S protein destabilization through receptor binding and proteolytic cleavage. These studies should also serve as a foundation for the structure-based design of betacoronavirus vaccine immunogens.**

Betacoronavirus S proteins are processed into S1 and S2 subunits by host proteases<sup>5</sup>. Like other class I viral fusion proteins, the two subunits trimerize and fold into a metastable pre-fusion conformation. The S1 subunit is responsible for receptor binding, while the S2 subunit mediates membrane fusion. Coronaviruses typically possess two domains within S1 capable of binding to host receptors: an amino (N)-terminal domain (NTD) and a carboxy (C)-terminal domain (CTD), with the latter recognizing protein receptors for SARS-CoV and MERS-CoV<sup>6,7</sup>. Although these individual domains have been structurally characterized, the organization of the complete spike has not yet been determined, preventing a mechanistic understanding of S protein function.

Here, we present the structure of the HKU1 S protein ectodomain determined using cryo-electron microscopy (cryo-EM) to 4.0 Å resolution (Fig. 1a and Extended Data Figs 1 and 2 and Extended Data Table 1). The protein construct contains a C-terminal T4 fibrin trimerization motif and a mutated S1/S2 furin-cleavage site (Extended Data Fig. 3). The S1 subunit adopts an extended conformation with short linkers between domains and sub-domains (Fig. 1b). The S1 NTD (amino acids 14–297) has strong structural and sequence homology to the bovine coronavirus (BCoV) S1 NTD (Extended Data Fig. 4), which recognizes acetylated sialic acids on glycosylated cell-surface receptors<sup>8</sup>. The glycan-binding site in the BCoV S1 NTD is conserved in the HKU1 S1 NTD and is located at the apex of the trimer, oriented towards target cells. Indeed, HKU1 S1 was recently shown to bind O-acetylated sialic

acids on host cells, and these glycans were required for efficient infection of primary human airway epithelial cultures<sup>9</sup>.

The HKU1 S1 CTD (amino acids 325–605) consists of a structurally conserved core connected to a large, variable loop (HKU1 S amino acids 428–587)<sup>10</sup> that is partially disordered (Extended Data Figs 5 and 6). The CTD is located at the trimer apex close to the threefold axis, and the core interacts with the other two S1 CTD cores and with one NTD from an adjacent protomer. The domain swapping between protomers results in a woven appearance when viewed looking down towards the viral membrane (Fig. 2a). Structural alignment of the SARS-CoV and MERS-CoV CTD–receptor complexes<sup>11,12</sup> with the HKU1 pre-fusion S protein reveals that the protein–receptor-binding surface of the S1 CTD is buried in the HKU1 S protein trimer and is therefore incapable of making equivalent interactions without some initial breathing and transient exposure of these domains (Fig. 2b).

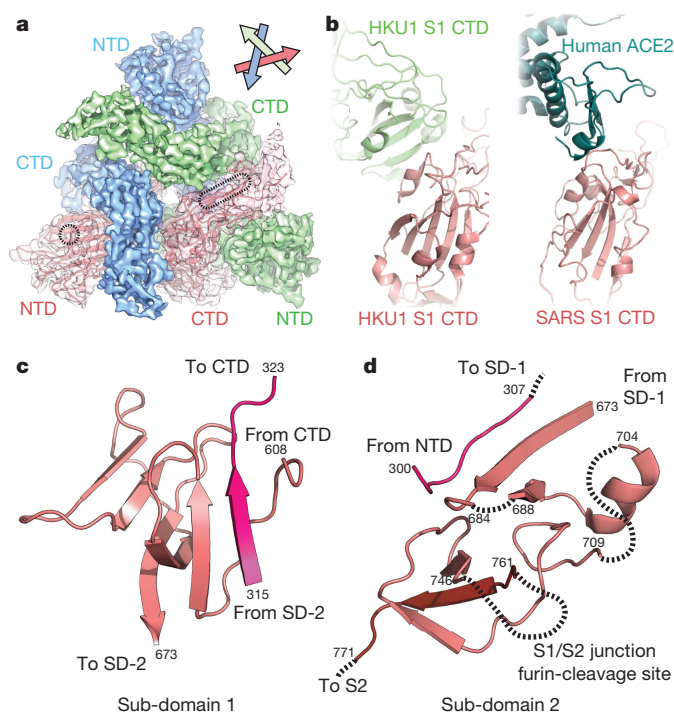


**Figure 1 | Structure of the HKU1 pre-fusion spike ectodomain.**

**a**, A single protomer of the trimeric S protein is shown in cartoon representation coloured as a rainbow from the N to C terminus (blue to red) with the reconstructed EM density of remaining protomers shown in white and grey. **b**, The S1 subunit is composed of the NTD and CTD as well as two sub-domains (SD-1 and SD-2). The S2 subunit contains the coronavirus fusion machinery and is primarily  $\alpha$ -helical. **c**, Domain architecture of the HKU1 S protein coloured as in **a**.

<sup>1</sup>Department of Integrative Structural and Computational Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>2</sup>Department of Biochemistry, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire 03755, USA. <sup>3</sup>Viral Pathogenesis Laboratory, National Institute of Allergy and Infectious Diseases, Building 40, Room 2502, 40 Convent Drive, Bethesda, Maryland 20892, USA. <sup>†</sup>Present address: Biomedical Research Center, Qatar University, QU-NRC, Zone 5, Room D130, Doha, Qatar.

\*These authors contributed equally to this work.



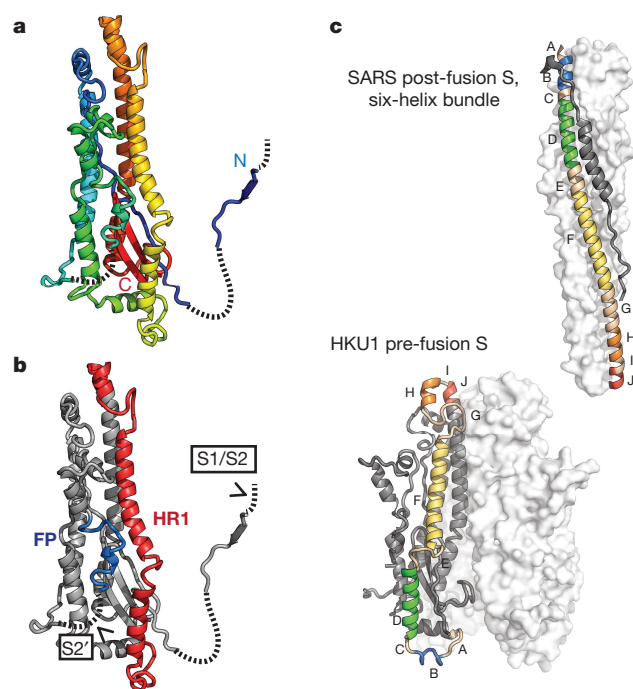
**Figure 2 | Architecture of the HKU1 S1 subunit.** **a**, EM density corresponding to each S1 protomer is shown. The putative glycan-binding and protein-receptor-binding sites are indicated with dashed shapes on the NTD and CTD, respectively. **b**, The HKU1 S1 CTD forms quaternary interactions with an adjacent CTD using a surface similar to that used by SARS CTD to bind its receptor, ACE2 (ref. 11). **c**, Sub-domain 1 is composed of amino acid residues before and after the S1 CTD. **d**, Sub-domain 2 is composed of S1 sequence C-terminal to the CTD, a short peptide following the NTD, and the N-terminal strand of S2, which follows the S1/S2 furin-cleavage site.

Although a protein receptor has not yet been identified for HKU1, antibodies against the CTD, but not those against the NTD, blocked HKU1 infection of cells<sup>13</sup>. These data suggest that the S1 CTD is the primary HKU1 receptor-binding site<sup>13</sup>, whereas the NTD mediates initial attachment via glycan binding.

HKU1 S1 also contains two sub-domains (which we term SD-1 and SD-2) that lack significant homology to previously determined structures (Fig. 2c, d). These sub-domains are primarily composed of S1 amino acid sequences following the CTD. However, stretches of amino acids preceding the CTD as well as S2 residues adjacent to the S1/S2 cleavage site also contribute to the sub-domains. This complex folding of elements dispersed throughout the primary sequence may allow receptor-induced conformational changes in the CTD to be transmitted to other parts of the structure.

In contrast to other viral fusion proteins such as influenza haemagglutinin (HA)<sup>14</sup> or HIV-1 envelope (Env)<sup>15,16</sup>, the HKU1 S1 subunits are rotated about the trimeric threefold axis with respect to the S2 subunits, causing the S1 subunit from one protomer to sit above the S2 subunit of an adjacent protomer (Extended Data Fig. 7). Similar to HA and Env, a region in the HKU1 S1 CTD (amino acids 371–380) caps the S2 central helix, thereby preventing the fusion machinery from springing into action.

Processing of coronavirus S proteins by host proteases plays a critical role in the entry process<sup>5</sup>. HKU1 S is cleaved by furin into S1 and S2 subunits during protein biosynthesis. Though mutated in the protein construct used here and disordered in the density map, the HKU1 S furin-cleavage site at the S1/S2 junction lies in a loop of SD-2 (Fig. 3 and Extended Data Fig. 6). Furin cleavage would leave a single S2  $\beta$ -strand participating in the SD-2  $\beta$ -sheets (Fig. 2d). Coronavirus S proteins also have a secondary cleavage site, termed S2' (Arg900)<sup>5</sup>,



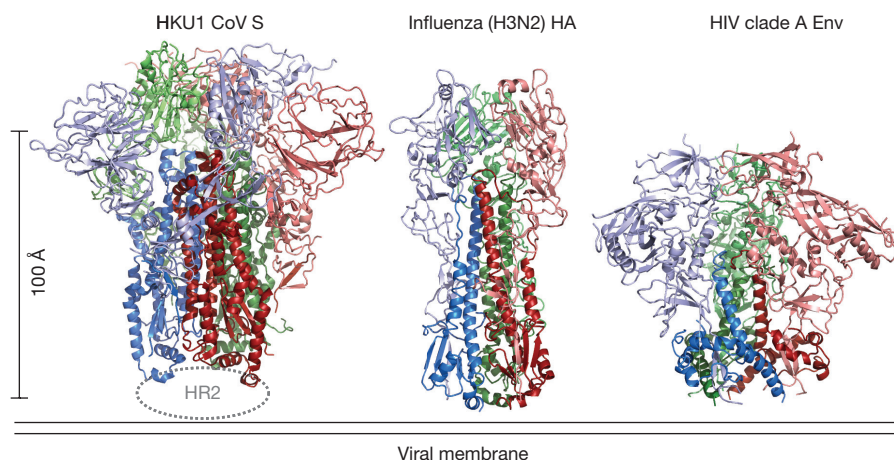
**Figure 3 | HKU1 S2 subunit fusion machinery.** **a**, The HKU1 S2 subunit is coloured like a rainbow from the N-terminal  $\beta$ -strand (blue), which participates in S1 sub-domain 2, to the C terminus (red) before HR2. **b**, The HKU1 S2 structure contains the fusion peptide (FP) and a heptad repeat (HR1). Protease-recognition sites are indicated within disordered regions of the protein (dashed lines). **c**, A comparison of coronavirus S2 HR1 in the pre- and post-fusion<sup>22</sup> conformations. Five HR1  $\alpha$ -helices are labelled and coloured like a rainbow from blue to red, N to C terminus, respectively. The structures are oriented to position similar portions of the central helix (red).

adjacent to the viral fusion peptide (amino acids 901–918)<sup>17</sup> (Fig. 3b and Extended Data Fig. 6). This is similar to the multiple endoproteolytic cleavage events that occur in the fusion proteins of respiratory syncytial virus (RSV) and Ebola virus<sup>18,19</sup>. Protease cleavage at S2' likely follows S1/S2 cleavage and may not occur until host-receptor engagement at the plasma membrane or viral endocytosis<sup>5</sup>.

As in all class I viral fusion proteins, the coronavirus S2 subunit contains the four elements required for membrane fusion: a fusion peptide or loop, two heptad repeats (HR1 and HR2), and a transmembrane domain<sup>14,20,21</sup>. Refolding of HR1 into a long  $\alpha$ -helix thrusts the fusion peptide into the host-cell membrane, and as the two heptad repeats interact to form a coiled-coil, the host and viral membranes are brought together. The fusion peptide, conserved among coronavirus S proteins<sup>17</sup> (Extended Data Fig. 6), is located on the exterior of the HKU1 S protein and is adjacent to the putative S2' cleavage site, which remains uncleaved in our structure. The fusion peptide forms a short helix and a loop, with most of the hydrophobic amino acids buried in an interface with other elements of S2. Unlike influenza HA where the C terminus of the fusion peptide is only 14 amino acids away from the N terminus of HR1, the fusion peptide of HKU1 S is 60 amino acids away from HR1. This span of protein contains four short  $\alpha$ -helices and several longer regions lacking regular secondary structure. This intervening sequence is also buried beneath SD-2 and the S2' cleavage site, suggesting that cleavage may affect the proclivity of S2 for undergoing the transition to the post-fusion conformation.

Coronavirus S protein heptad repeats are unusually large with HR1 encompassing more than 90 amino acids<sup>20</sup>. In the cryo-EM structure, HR2 is located at the base of the HKU1 S protein near the viral membrane, but is poorly ordered, precluding unambiguous assignment of the residues. However, HR1 is well ordered and arranged along the length of the S2 subunit, forming four short helices and part of the





**Figure 4 | Comparison of structurally related class I viral fusion proteins.** The fusion proteins from coronaviruses, influenza virus and HIV-1 are cleaved into receptor-binding subunits (pink, light green, light blue) and the viral fusion machinery (dark red, dark green, blue)<sup>14–16,28</sup>. Comparison to other class I fusion proteins can be found in Extended Data Fig. 8.

central three-helix bundle. This arrangement of HR1 is similar to that of influenza HA, although in HA the HR1 is organized as two helices connected by a long loop<sup>14</sup>. Conversion of influenza HA to the post-fusion conformation requires these protein elements to transition into a single long  $\alpha$ -helix<sup>21</sup>. The post-fusion six-helix bundle structures of SARS-CoV and MERS-CoV S2 heptad repeats<sup>22,23</sup> reveal that coronavirus S proteins also undergo a similar transition (Fig. 3c). However, the S protein must carry out five such loop-to-helix transitions, highlighting the complexity of S proteins relative to other class I fusion proteins. In addition, the membrane distal regions of the pre-fusion S2 central three-helix bundle (S2 amino acids 1070–1076), which is the C-terminal portion of HR1, are splayed outwards from the threefold axis (Extended Data Fig. 7). In the available coronavirus post-fusion HR1–HR2 structures, this portion of HR1 forms a tight three-helix bundle<sup>22,23</sup>. Formation of this three-helix bundle may be prevented by interactions between the C-terminal end of the S2 HR1 and the S1 CTD, and thus disruption of these interactions through receptor-induced conformational changes would provide an additional means by which receptor binding in S1 can initiate S2-mediated membrane fusion. Indeed, protease cleavage and an acidic pH are thought to be insufficient to trigger the transition to the post-fusion conformation without additional destabilization provided by receptor binding<sup>24–26</sup>.

The formation of anti-parallel six-helix bundles composed of HR1 and HR2 in the post-fusion conformation is a unifying feature of class I viral fusion proteins. However, the pre-fusion conformations of this protein family are incredibly diverse in size and topology (Extended Data Fig. 8). The HKU1 S protein structure presented here most closely resembles influenza virus HA and HIV-1 Env (Fig. 4), which also have receptor-binding subunits that cap the central helix of the fusion subunit<sup>14,15,27,28</sup>. However, some core elements of the fusion machinery are conserved amongst all class I fusion proteins, including paramyxovirus F proteins.

The HCoV-HKU1 S protein trimer in a pre-fusion conformation is, to our knowledge, the largest class I viral fusion glycoprotein structure determined to date (Fig. 4 and Extended Data Figs 8 and 9). Since betacoronavirus S proteins are similar in size and have a conserved domain organization, our findings should be generally applicable to other betacoronaviruses, including SARS-CoV and MERS-CoV (Extended Data Fig. 6). Our studies provide a structural basis for S protein function wherein the pre-fusion S protein is progressively matured and destabilized by receptor binding and protease cleavage. Following dissociation of the S1 subunits, HR1 would transition to a long  $\alpha$ -helix, and the fusion peptide would be released from the side of the S2 subunit and inserted into host membranes. The structure and mechanistic insights presented here should enable engineering of pre-fusion

stabilized coronavirus S proteins as vaccine immunogens against current and emerging betacoronaviruses, similar to recent efforts for other viral fusion proteins<sup>29,30</sup>. This work also acts as a springboard for future studies to define mechanisms of antibody recognition and neutralization, which will lead to an improved understanding of coronavirus immunity.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 18 January; accepted 5 February 2016.**

- Woo, P. C. *et al.* Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* **79**, 884–895 (2005).
- Christian, M. D., Poutanen, S. M., Loutfy, M. R., Muller, M. P. & Low, D. E. Severe acute respiratory syndrome. *Clin. Infect. Dis.* **38**, 1420–1427 (2004).
- Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
- Graham, R. L. & Baric, R. S. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J. Virol.* **84**, 3134–3146 (2010).
- Millet, J. K. & Whittaker, G. R. Host cell proteases: critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* **202**, 120–134 (2015).
- Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).
- Mou, H. *et al.* The receptor binding domain of the new Middle East respiratory syndrome coronavirus maps to a 231-residue region in the spike protein that efficiently elicits neutralizing antibodies. *J. Virol.* **87**, 9379–9383 (2013).
- Peng, G. *et al.* Crystal structure of bovine coronavirus spike protein lectin domain. *J. Biol. Chem.* **287**, 41931–41938 (2012).
- Huang, X. *et al.* Human coronavirus HKU1 spike protein uses O-acetylated sialic acid as an attachment receptor determinant and employs hemagglutinin-esterase protein as a receptor-destroying enzyme. *J. Virol.* **89**, 7202–7213 (2015).
- Li, F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J. Virol.* **89**, 1954–1964 (2015).
- Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
- Lu, G. *et al.* Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* **500**, 227–231 (2013).
- Qian, Z. *et al.* Identification of the receptor-binding domain of the spike glycoprotein of human betacoronavirus HKU1. *J. Virol.* **89**, 8816–8827 (2015).
- Wilson, I. A., Skehel, J. J. & Wiley, D. C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373 (1981).
- Julien, J. P. *et al.* Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1477–1483 (2013).
- Lyumkis, D. *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1484–1490 (2013).
- Madu, I. G., Roth, S. L., Belouzard, S. & Whittaker, G. R. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J. Virol.* **83**, 7411–7421 (2009).

18. Schornberg, K. *et al.* Role of endosomal cathepsins in entry mediated by the Ebola virus glycoprotein. *J. Virol.* **80**, 4174–4178 (2006).
19. Zimmer, G., Budz, L. & Herrler, G. Proteolytic activation of respiratory syncytial virus fusion protein. Cleavage at two furin consensus sequences. *J. Biol. Chem.* **276**, 31642–31650 (2001).
20. Bosch, B. J., van der Zee, R., de Haan, C. A. & Rottier, P. J. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* **77**, 8801–8811 (2003).
21. Bullough, P. A., Hughson, F. M., Skehel, J. J. & Wiley, D. C. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43 (1994).
22. Duquerroy, S., Vigouroux, A., Rottier, P. J., Rey, F. A. & Bosch, B. J. Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein. *Virology* **335**, 276–285 (2005).
23. Lu, L. *et al.* Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nature Commun.* **5**, 3067 (2014).
24. Li, F. *et al.* Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. *J. Virol.* **80**, 6794–6800 (2006).
25. Matsuyama, S. & Taguchi, F. Two-step conformational changes in a coronavirus envelope glycoprotein mediated by receptor binding and proteolysis. *J. Virol.* **83**, 11133–11141 (2009).
26. Simmons, G. *et al.* Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc. Natl Acad. Sci. USA* **102**, 11876–11881 (2005).
27. Kong, L. *et al.* Complete epitopes for vaccine design derived from a crystal structure of the broadly neutralizing antibodies PGT128 and 8ANC195 in complex with an HIV-1 Env trimer. *Acta Crystallogr. D* **71**, 2099–2108 (2015).
28. Lee, P. S. *et al.* Receptor mimicry by antibody F045–092 facilitates universal binding to the H3 subtype of influenza virus. *Nature Commun.* **5**, 3614 (2014).
29. McLellan, J. S. *et al.* Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* **342**, 592–598 (2013).
30. Sanders, R. W. *et al.* A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLoS Pathog.* **9**, e1003618 (2013).

**Acknowledgements** We thank colleagues and members of our laboratories for critical reading of the manuscript. This work was supported by start-up funds to A.B.W. from the Scripps Research Institute and to J.S.M. from the Geisel School of Medicine at Dartmouth, and funding from intramural NIAID to B.S.G. Partial salary support for R.N.K. was provided by NIH grant R56 AI118016. This is manuscript 29286 from TSRI.

**Author Contributions** H.M.Y. and K.S.C. designed the HKU1 S expression constructs. N.W. expressed and purified the proteins. C.A.C., H.L.T., and J.P. prepared samples for cryo-EM, collected images, and processed the data. R.N.K. and C.A.C. built and refined the model. B.S.G., J.S.M., and A.B.W. conceived of the project. R.N.K., C.A.C., N.W., B.S.G., J.S.M. and A.B.W. analysed the results and wrote the manuscript, with all authors editing and approving the final manuscript.

**Author Information** The negative stain EM maps of the HKU1 spike ectodomain have been deposited in the EMDB under accession numbers EMD-8066, EMD-8067, and EMD-8068. The cryo-EM map of the HKU1 spike ectodomain has been deposited in the EMDB under accession number EMD-8069 and the coordinates of the HKU1 spike ectodomain structure have been deposited in the PDB under accession code 5I08. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S.M. (Jason.S.McLellan@Dartmouth.edu) or A.B.W. (ABWard@Scripps.edu).

## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

**Protein expression and purification.** A mammalian-codon-optimized gene encoding HKU1 S (isolate N5, NCBI accession Q0ZME7) residues 1–1276 with a C-terminal T4 fibrin trimerization domain, a HRV3C cleavage site, and a 6xHis-tag was synthesized and subcloned into the eukaryotic expression vector pVRC8400. The S1/S2 furin-recognition site 752-RRKRR-756 was mutated to GSGS to generate the uncleaved construct used for cryoEM studies. Three hours after this plasmid was transfected into FreeStyle 293-F cells (Invitrogen), kifunensine was added to a final concentration of 5  $\mu$ M. FreeStyle 293-F cells are a high-transfection-efficiency cell line adapted for suspension culture derived from low passage clonal cultures and after purchase were not further authenticated. Cells were not confirmed to be free of mycoplasma, but were only used for protein expression. Cultures were harvested after six days, and protein was purified from the medium using Ni-NTA Superflow resin (Qiagen). The buffer was then exchanged using a HiPrep 26/10 desalting column (GE Healthcare Biosciences) from a high-imidazole elution buffer to a low pH buffer (20 mM Bis-Tris pH 6.5, 150 mM NaCl). Afterward, endoglycosidase H (EndoH) (10% w/w) and HRV3C protease (1% w/w) were added to the protein and the reaction was incubated overnight at 4°C. The digested protein was further purified using a Superose 6 16/70 column (GE Healthcare Biosciences).

The furin-cleaved HKU1 S construct analysed by negative-stain EM was similar to the one described above except that it encoded residues 1–1249 and contained the wild-type RRKRR furin-recognition site. Expression and purification were also similar, except that a plasmid expressing furin was co-transfected into the FreeStyle 293-F cells to ensure complete processing of the protein.

**Sample preparation for negative-stain electron microscopy.** HKU1 S proteins were placed directly onto 400 copper mesh grids and then stained with 1% uranyl formate. Tris-buffered saline (TBS) was used as buffer if dilution was necessary.

**Negative-stain electron microscopy data collection.** Grids were loaded into a Tecnai T12 Spirit operating at 120 keV and imaged using a Tietz TemCam-F416 CMOS at 52,000 $\times$  magnification at  $\sim$ 1.5  $\mu$ m under focus. Micrographs were collected using Leginon<sup>31</sup> and processed within Appion<sup>32</sup>. Particles were picked using a difference-of-Gaussians approach<sup>33</sup> and aligned using reference-free 2D classification employing iterative multivariate statistical analysis/multi-reference alignment (MRA/MSA) using a binning factor of 2 to remove amorphous particles<sup>34</sup>. Particles in classes that did not represent views of HKU1 S proteins were discarded. ISAC<sup>35</sup> was used to generate a template stack from which initial 3D models were generated using the EMAN2 (ref. 36) procedure initialmodel.py. 3D models were refined using EMAN1 (ref. 37).

**Sample preparation for cryo-electron microscopy.** Sample solution (3  $\mu$ l) was applied to the carbon face of a CF-2/2-4C C-Flat grid (Electron Microscopy Sciences, Protochips) that had been plasma cleaned for five seconds using a mixture of Ar/O<sub>2</sub> (Gatan Solarus 950 Plasma system). The grid was then manually blotted and immediately plunged into liquid ethane using a manual freeze plunger.

**Cryo-electron microscopy data collection.** Movies were collected via the Leginon interface on a FEI Titan Krios operating at 300 keV mounted with a Gatan K2 direct-electron detector<sup>31</sup>. Each movie was collected in counting mode at 22,500 $\times$  nominal magnification resulting in a calibrated pixel size of 1.31 Å/pix at the object level. A dose rate of  $\sim$ 10 e<sup>−</sup>/((cam pix)  $\times$  s) was used; exposure time was 200 ms per frame. The data collection resulted in a total of 1,049 movies containing 50 frames each. Total dose per movie was 57 e<sup>−</sup>/Å<sup>2</sup>. Data were collected at 1.0 to 3.5  $\mu$ m under focus.

**Cryo-electron microscopy data processing.** Frames in each movie were aligned<sup>38</sup>, and CTF estimation was carried out using CTFFIND3 (ref. 39). Particles were picked from a subset of the data employing a difference-of-Gaussians approach<sup>33</sup> and aligned using reference-free 2D classification employing iterative MRA/MSA using a binning factor of two<sup>34</sup>. The resulting 2,188 particles were used to generate an initial 25 Å lowpass-filtered 3D reconstruction using EMAN2. SPIDER refproj.spt<sup>40</sup> with a delta theta angle of 15 degrees was used to generate 83 projection images of the initial 3D reconstruction. These projection images were used as templates for picking particles from the entire cryo data set. Particles from the entire data set were aligned and classified with the same methods used for the subset of particles stated above. After 2D classification, unbinned selected particles were symmetrically refined in RELION version 1.3 (refs 41, 42) against the initial 3D reconstruction filtered to 60 Å resolution. This refinement was followed by particle polishing and refinement of the resulting realigned, B-factor-weighted and signal-integrated particles using RELION version 1.4b1. The resolution of the final map was 4.04 Å at an FSC cutoff of 0.143. A mask was generated in RELION using a threshold that accounted for the entire structure. From this threshold, the mask was further dilated by 3 voxels and a Gaussian fall-off was generated over an

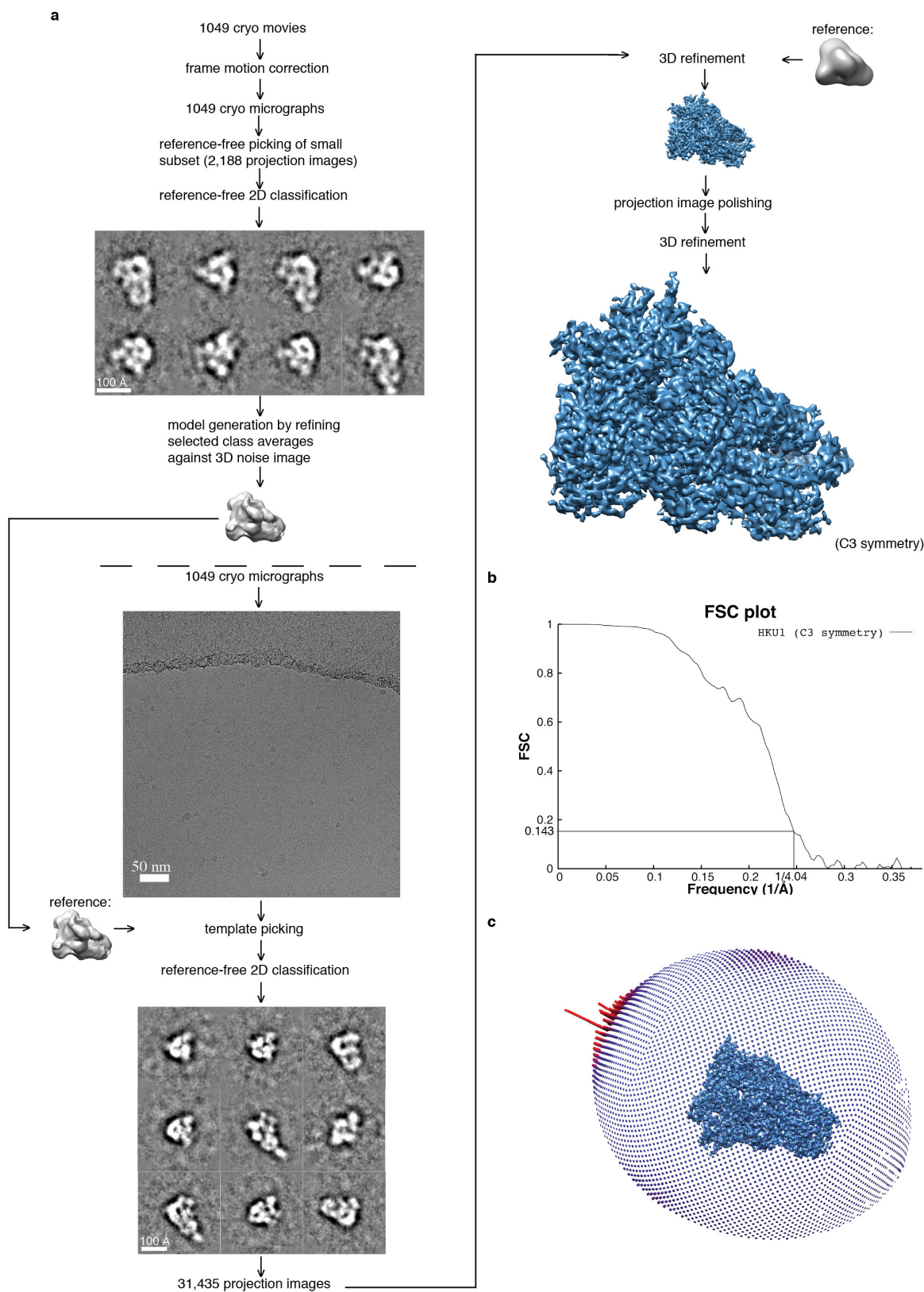
additional 6 voxels. The mask effect on FSC was taken into consideration. Phases were randomized in the unfiltered half-set maps for initial FSC lower than 0.8 and a new FSC between these phase-randomized maps was generated and used to correct for mask effects in the final FSC-based resolution estimate. The reported resolution of 4.04 Å is the RELION CorrelationCorrected value.

The map was B-factor sharpened employing FSC-weighting. The B-factor was estimated in RELION based on the resolution range from 10 Å to 2.62 Å (B-factor =  $-117 \text{ Å}^2$ ). The detector MTF file was provided to RELION.

**Model building and refinement.** An initial model of the S1 NTD was generated using the Modeller<sup>43</sup> homology modelling tool in UCSF Chimera<sup>44</sup> with the BCoV NTD (PDB 4H14)<sup>8</sup> as a template. The NTD homology model was docked into the HKU1 S protein EM density and refined with Rosetta density-guided iterative local refinement<sup>45</sup> while imposing C3 symmetry. Rosetta output models were clustered based on pairwise r.m.s.d. using a cluster radius of 2.15 Å. The lowest energy model from the largest cluster was selected for additional refinement. This model and the conserved CTD core from SARS-CoV (PDB 2AJF)<sup>11</sup> were used as starting structures for model building and refinement. These starting models and the remaining HKU1 protein sequence were modelled manually using COOT<sup>46</sup> and refined using RosettaRelax<sup>47</sup>. Structures were evaluated using EMRinger<sup>48</sup> and Molprobity<sup>49</sup>. Figures were produced in the PyMol<sup>50</sup> or UCSF Chimera<sup>44</sup> software packages.

31. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
32. Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
33. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).
34. Ogura, T., Iwasaki, K. & Sato, C. Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking. *J. Struct. Biol.* **143**, 185–200 (2003).
35. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20**, 237–247 (2012).
36. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
37. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
38. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
39. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
40. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
41. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
42. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
43. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **47**, 5.6.1–5.6.32 (2014).
44. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
45. DiMaio, F. *et al.* Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nature Methods* **12**, 361–365 (2015).
46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
47. DiMaio, F. *et al.* Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
48. Barad, B. A. *et al.* EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nature Methods* **12**, 943–946 (2015).
49. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
50. Schrödinger, L. The PyMOL Molecular Graphics System, Version 1.5.0.4. (2012).
51. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
52. Sievers, F. & Higgins, D. G. Clustal Omega. *Curr. Protoc. Bioinformatics* **48**, 13.13.11–13.13.16 (2014).
53. Lee, J. E. *et al.* Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454**, 177–182 (2008).
54. McLellan, J. S. *et al.* Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**, 1113–1117 (2013).
55. Welch, B. D. *et al.* Structure of the cleavage-activated prefusion form of the parainfluenza virus 5 fusion protein. *Proc. Natl Acad. Sci. USA* **109**, 16672–16677 (2012).
56. Wong, J. J., Paterson, R. G., Lamb, R. A. & Jardetzky, T. S. Structure and stabilization of the Hendra virus F glycoprotein in its prefusion form. *Proc. Natl Acad. Sci. USA* **113**, 1056–1061 (2016).
57. Xu, K. *et al.* Crystal structure of the pre-fusion Nipah virus fusion glycoprotein reveals a novel hexamer-of-trimers assembly. *PLoS Pathog.* **11**, e1005322 (2015).

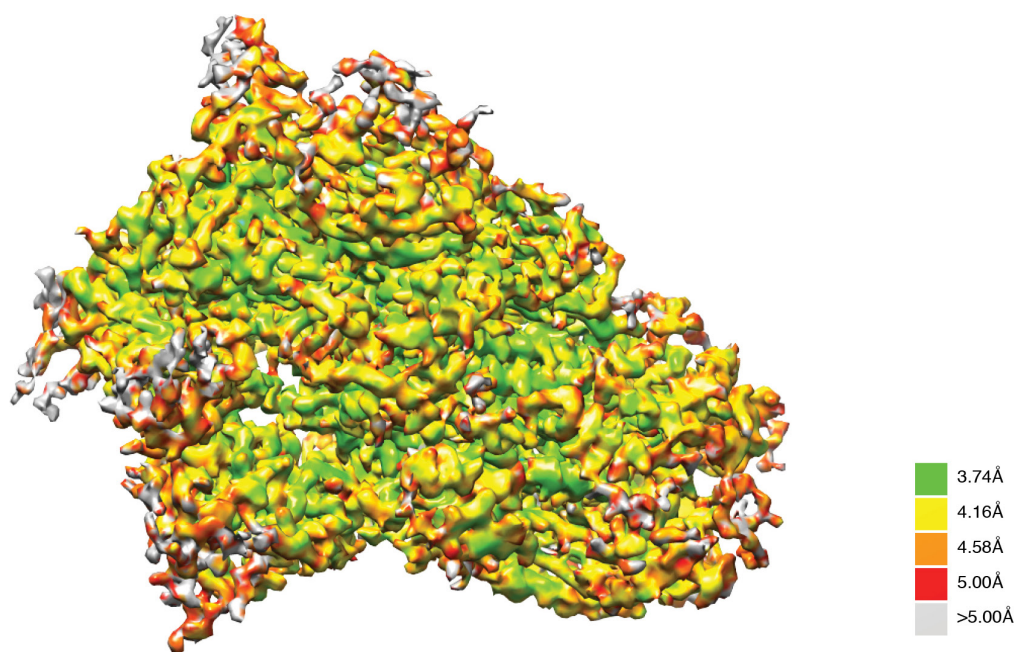




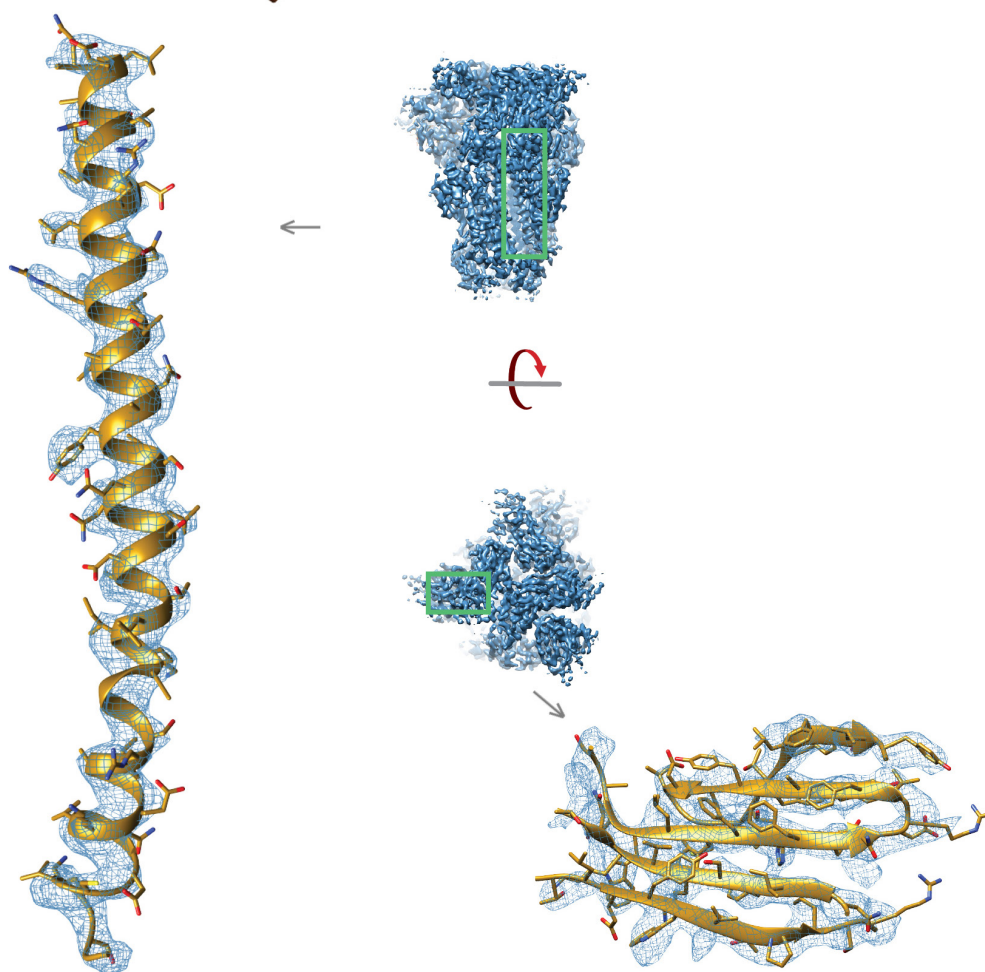
**Extended Data Figure 1 | Data processing flowchart.** **a**, Processing resulting in density map of pre-fusion HKU1 spike glycoprotein at 4.04 Å resolution. **b**, FSC plot illustrating correlation between two volumes refined independently from two distinct half sets of raw data. A final

resolution of 4.04 Å is indicated in the plot. **c**, Angular distribution of raw data within the data set. A slight, but within normal range, over-representation of top views was observed (tall red bars).

a

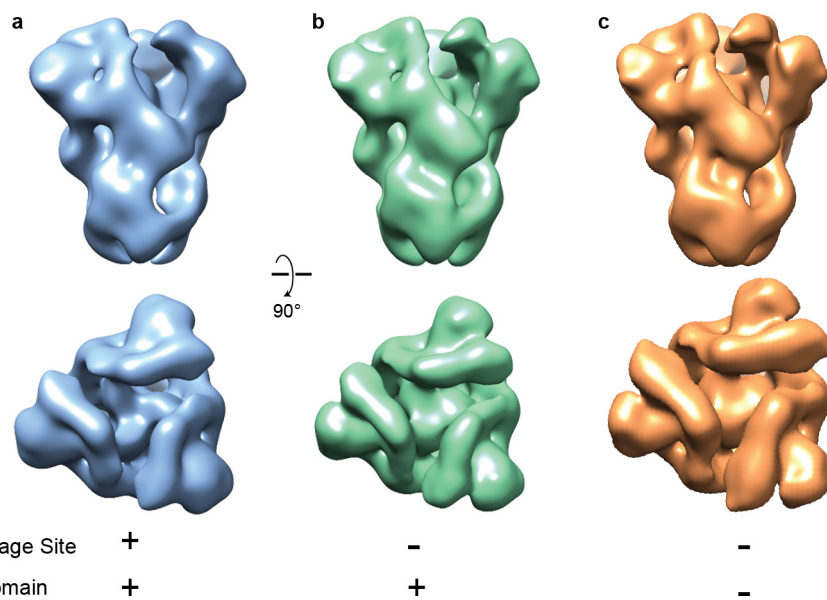


b



**Extended Data Figure 2 | Resolution of the pre-fusion HKU1 S density map.** **a**, Local resolution within the EM density map. Local resolution was calculated using ResMap<sup>51</sup> discretizing every 0.25 Å over a range from  $2 \times$  voxel size (2.62 Å) to  $4 \times$  voxel size (5.24 Å). Resolution significance criterion was set to 0.05. The resolution ranges from 3.74 Å

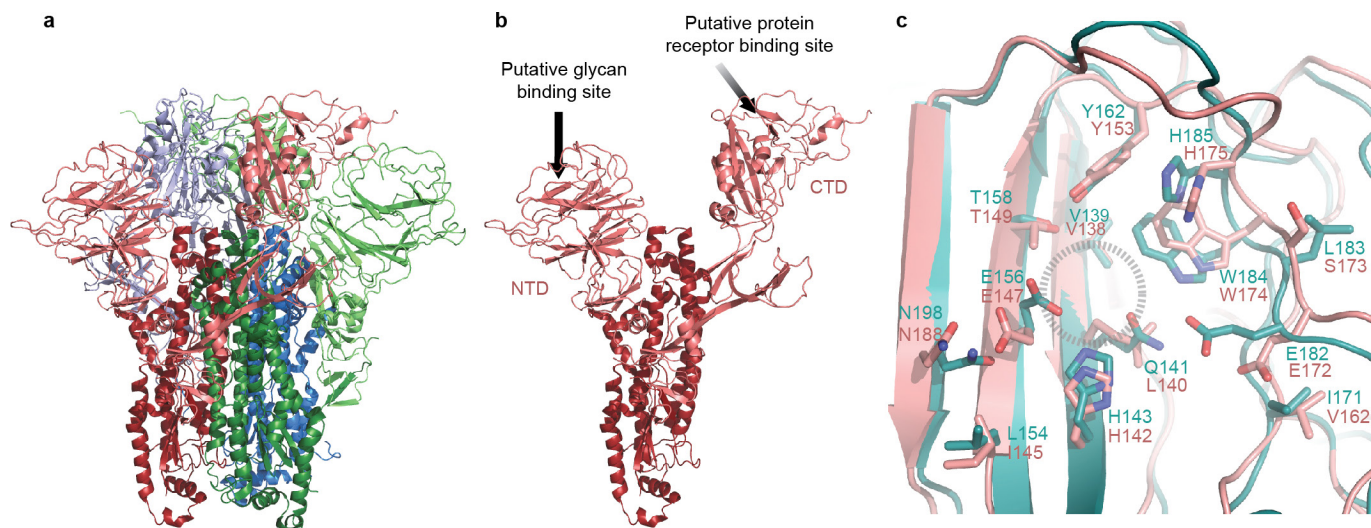
in stable internal secondary structures to greater than 5.00 Å in flexible peripheral loops. **b**, Close-ups of secondary-structure densities. To the left is displayed the central  $\alpha$ -helix of an S2 monomer and to the right is a  $\beta$ -sheet from the NTD domain in an S1 monomer.



**Extended Data Figure 3 | Cleavage at the S1/S2 junction does not induce large conformational changes in HKU1 spike.** **a**, HKU1 spike 1–1249 with an attached foldon domain and wild-type furin-cleavage site was reconstructed using negative-stain electron microscopy. **b**, HKU1

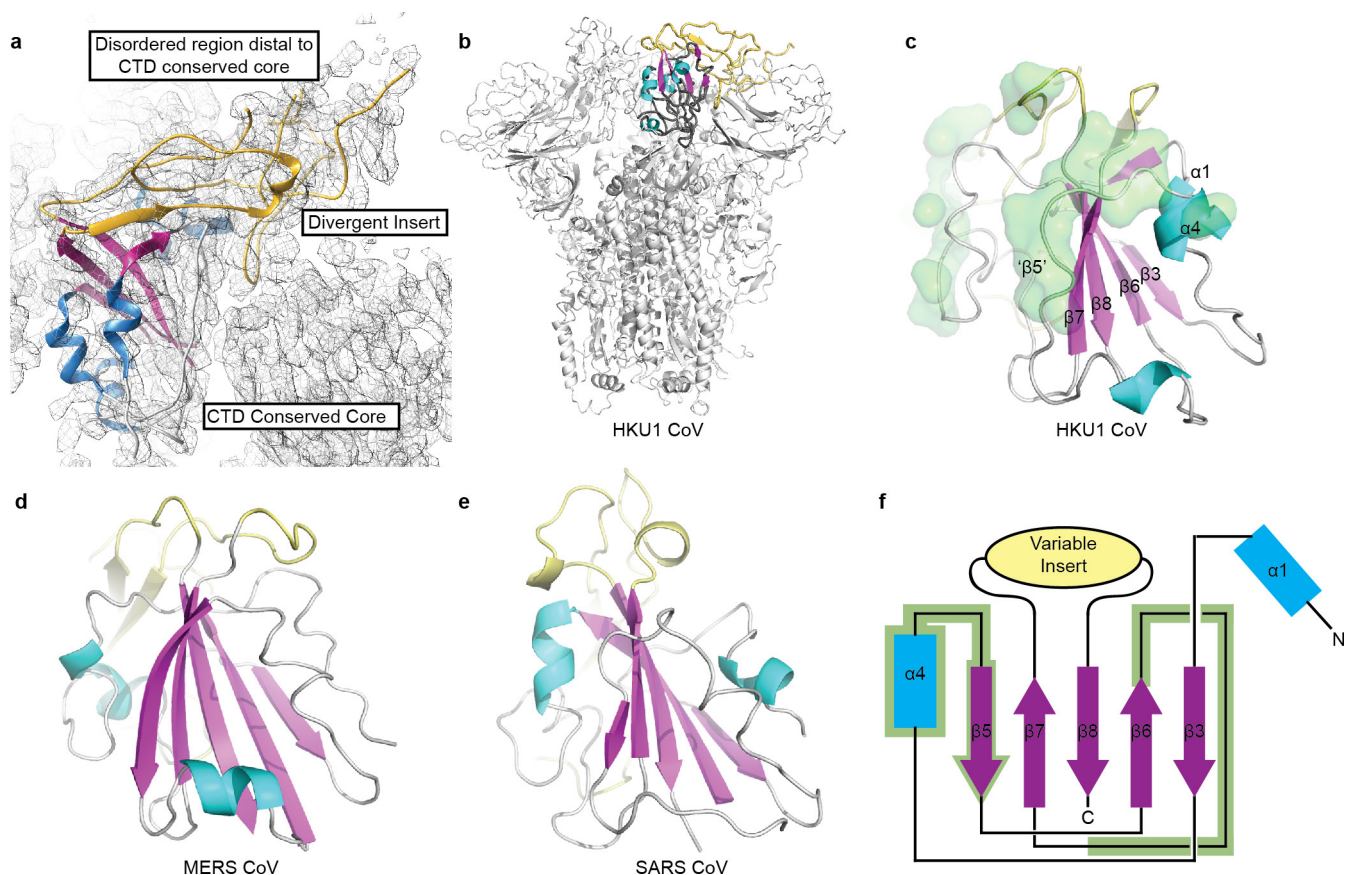
spike 1–1276 with an attached foldon and a mutated furin-cleavage site reconstructed using negative-stain electron microscopy. **c**, HKU1 spike 1–1249 without foldon and with mutated furin-cleavage site. Side and top views are shown.





**Extended Data Figure 4 | Putative glycan binding site of the HKU1 S1 NTD.** **a**, HKU1 trimeric S and **b**, an isolated monomer. Putative host glycan-binding and protein-receptor-binding sites are indicated. **c**, The bovine coronavirus (BCoV) S1 NTD structure from Peng *et al.*<sup>8</sup> (teal) is superposed onto the HKU1 S NTD (pink). Residue side-chains

involved in the putative glycan-binding site (dashed circle) are shown as sticks, with oxygen atoms coloured red and nitrogen atoms coloured blue. Note that N198 (BCoV) and N188 (HKU1) are predicted *N*-linked glycosylation sites.



**Extended Data Figure 5 | Betacoronavirus S proteins possess a conserved structural core in their C-terminal domains.** **a**, The structurally divergent loop of the S1 CTD is poorly ordered distal to the core CTD domain. The conserved S1 CTD cores<sup>10</sup> of **b**, HKU1-CoV highlighted in the trimeric pre-fusion S, **c**, HKU1-CoV as an isolated domain, **d**, MERS-CoV<sup>12</sup> and **e**, SARS-CoV<sup>11</sup> are coloured according to secondary structure (β-sheets: pink, α-helices: blue, lacking regular

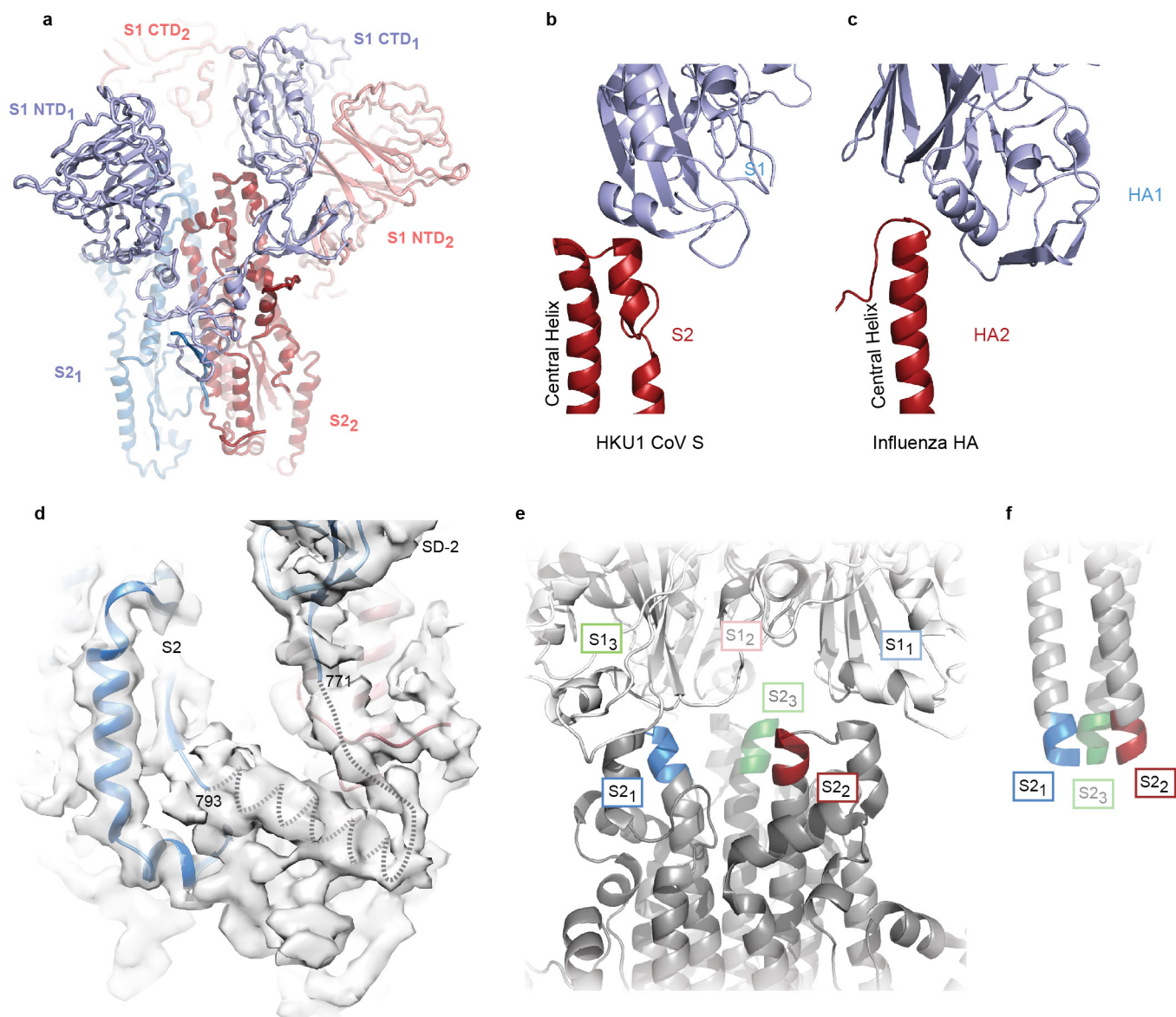
secondary structure: grey) and the insert which differs amongst coronaviruses is coloured yellow. Atoms participating in quaternary interactions with other HKU1 S protomer CTDs are shown in green surface in **c**, **f**. The positions of these interacting atoms are mapped on to the conserved core topology. The sheet and helix nomenclature is taken from reference 10.

		NTD																
HKU1	----	MFLIIFILPTTLAV--IGD-----	FNCTNSF--INDYNTKIPRISEDDVVDVSLGLTGYVVLNRVYLNNTLLFTGYFPKSGANFRDLALKGSIY-----	LSTLWYKPPFLSDFNN	100													
SARS	----	MFIFLLFLTLTS-----	G--SDLCRCTTFD-----	DVQAPNYT--QHTSSMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGFHT-----	INHTEGN-PVIPFKD	85												
MERS	MIHSVFLMLFLTPTES	XVDVGSDSVKSACIEVDLQQTFDFDCTWFR-P----	IDVSKADGIIYPGRTYSNITITYQG-LFFYQGDHGMVYVSAGHATGTTTPQKFLVANSYQ-DVKQFAN		114													
		:*:::.*	*	*	:	*	..*	*	*	:	:	..:	:	:	:	*	:	
HKU1	GIFSKVNKTLYVNN-----	TLYSEFSTIVIGSVFVNTSYTIV--	VQPHNGILEITACQYTMCEYPHTVCKSKGSIRNE-----	SWHDSSE-PLCLFKNFTYVNSAD	196													
SARS	GIYFAATEK-----	SNVVRGWVFGSTMNKSQSVI--	IINNSTNVIRACNFELCDNPFPAVSKPMGTQHTMIFDNAFNCTFEYISDAF--	SLDSEKSGNFKHL--	182													
MERS	GFVVRIGAAANSTGTVII	SPSTSATIRKIYPAFMLGSSVGNFSDGKMGFRFNHTLVLLPDGCGTL--	LRAYCYILEPRSGNHCP--	AGNSYTSFATYHTPATDCSDGNYNRNASLNSF--	228													
	*	:	..*	*	:	..:	:	:	:	:	:	:	:	:	:	:	:	:
HKU1	WLYFHFYQERGIFYAYY-----	ADV-GMPTTFLFSLYLGTILSHYYVMPLTCNAISSNTD---	NETLEYWVTPLSRRQYLLNFDEHGVTINAVDCSSSF		286													
SARS	-REFVFNKNGDGLYVYKGYQPID--	VVRDLPSGFGNTLKP-----	IFKLPLGINITNFRAI--	LTAFSPAQDIWGTSAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP	282													
MERS	-KEYFNLNRCNCTFMYTINTEDEILEWFGITQTAQGVHLFSSRYVDLYGGMFQFATLPYDITIKYYSIIPHSIRSIQSDRKAW---	AAFYVYKQLPLTFLDLFSDGVIIRRAIDCGFND			343													
	:	..:	..*	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
			CTD															
HKU1	LSEIQCKTQSFAPNTGVYDLSGFTVKPVATVYRRIPNLPC	DCDIDNWNLVNVPSPNLNWRERIFSNCFNLSTLLRLVHVDVSCNNLDKSKIFGSCFNSITVDKFAIPNRNRDDLQLGSS			406													
SARS	LAEIKCSVKSEIDKGIYQTSNFRVVPVGGDVVRFNPTNL	CPFGGEVFNATKFPVSVYAWERKISNCVADYVLYNSTFFSTFKCYGVSAATKLNLDLCSFNIVYADSVFVKGDDVRQIAPGQT			402													
MERS	LSQLHCSYESFDVBSGVSVSSFEAKPSGVSVEQAEG-VECD	FSPLLS-GTPPQVYNFKRLVETNCNLYNLKLLSLFVNDFTCQSQISPAIASNCYSSLLIDYFSYPLSMKSLDGLVSSA			461													
	*:::.*	::*:	::*:	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
			Variable Loop															
HKU1	GFLQSSNYKIDISSSSCQLYSL	LPVNVITINNFPSSWNRRYFGGSFNLSSYDVVYSDHCFVSNS-----	DFCPCADPSVNSCAKSKPPSAICPAGTKYRHCDDLTTLYVK		513													
SARS	GVIADYNYKLDDFMGCVLAWNTRNIDATSTGN-----	YNYKY-----	RYLRHG--KLRFERDISNVFSPDGKPTCPA-----	LNCYWFLNDYG-----	482													
MERS	GPISQVNYKQSFNSPTCLILATVPHNLITITKP-----	LKY-----	SYNKCRRLLSDRTEVPQLVNAVQYSPCVSIV-----	PSTVWEDGDYYRKQL--SPLEGG	551													
	*	:	..*	*	:	:	:	:	:	:	:	:	:	:	:	:	:	:
HKU1	NWCRCSCLPDP	ISTYSPTCPQKKVVVIGEGHCPGLGINEEKCQTQLNHSSCFCS	PDFAFLGWSFDCISNNRCNIFSNFIFNGINS	SGTTCSDLL--YSNTEISTGVCVNYDLYGITQGQ	631													
SARS	-FYTTTG-----	IG--YQYRVVV-----	LSFELL---	NAPATVCGP-----	KLSTDLIKNCQVNFNGLTGTG	536												
MERS	GWLVASG-----	ST--VAMTEQLQ-----	MGFGITVQYGTDTNSVCPKLEFANDTKIASQGLNCVYS	SLYVSGRG	615													
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
HKU1	IFKEVSAAYYNNQNL	LYDSNGNIIGF----	KDFTLNKTYTILPCYSGRVSAAFY--	QNSSPALLYRNLCYSYLVNNSIFISQ-----	PFYFDSYLGCVLNAVNLTSYSVS	732												
SARS	VLTSSKRF-QPFQQFGRD-----	VSDFTDSVRDPKTS	ELDISPCAFGGSVITPTGNASSEVAVLYQDVNCTDVSTAIHADQLT--	PAWRIYSTGMNVFQTQAGCLIGAEHVD--	TSY	646												
MERS	VFQNCITAVG-VRQRFVYDAYQNL	VGYSYD-----	DGNYCYCLACVSVFVSVIYD--	KETKTHATLFGSVACHEISSTMSQYSRSTRSMLKRRDSTYGLPQ	YPVGVGLVGNSS-LFVE	725												
	::	:	:	*	:	:	*	::	*	*	*	*	*	*	*	*	*	*
			S1/S2 <sup>v</sup>	S2														
HKU1	SCDLRMGSGFCIDYALPSSRRKR	RGISPPYR--FVTFEFNVSVFVNDVETVGG	LFEIQIPTNFTIAGHEEFITQSSPKVITDCSAFVCSNYAACHDLLSEYGTFC	DNINSILNEVDLL	850													
SARS	ECDIPIGAGICASYHTVSL	LRSTQSKSIVA-----	YTMSLGA-DSSIAYSNNITAIPTNFSISITTEVMPVSMARKTSVDCNMYICGDS	TECANLLQYGSFCTQLNRALSIGIAEQ	756													
MERS	DCKLPLGQSLCALPDP	TPSTLTPRSVRVPGEMRLASI-AFNHPIQV--	DQFNSSYFKLSIPTNFSFGVGTQYEQTTIQKVTVDCKQYICNGFQKCEQL	LRREYQGFCSKINQALHGANLRQ	842													
	..:	:	..*	*	:	:	:	:	:	:	:	:	:	:	:	:	:	:
			S2 <sup>v</sup>	FP														
HKU1	DITQLQVANALMGQVLTSSNL	NLNLHSDVDNIDFKSLGCLGSGCQSGSSR	SLLEDLLFNKVKLS	SDVGFVEAYNNCT--	GGSEIRDLLCVQSFNGIKVLPILSETQISGYTTAATVAAMF	968												
SARS	DRNTREVAQVKQMY-----	KPTTLKYFGG-FNFSQILPDPL--	KPTKRSFIEDLLFNKVT	LADAGFMKQYGECL--	GDINARDLICAKFNGLTVLPPLTDDMIAAYTAALVSGTAT	865												
MERS	DDSVRLNFASVKSQ-----	SSPIIPGEGDGNLTLLPEVSISTGSR	SARSIAEDLLFDKVTIADPGV	MQGYDDCQQGPASARDLICAQYVAGYKVLPLMDVNMEAA	YTSLLGSIAG	957												
	*	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
			HR1															
HKU1	PPWSA----	AAGVPFSLNVQYR	INGLVGTMVNLNKNQKLIANAFNKALLSIQNGFTATNSALAKIQSVVNANAQALNSLLQQLFNKFGA	ISSSLQEILSRDLNLEAQVQIDRLINGRLTA	1084													
SARS	AGWTFGAGAA	LPIPFAMQAYR	FNIGVGTQNVLYENQKQIANQFNKAI	SIQIQESLTTTSTALGKLQDVVNQAQALNTLVKQLSSNFGA	ISSVLNLDLSRLDKVEAEVQIDRLITGRLQS	985												
MERS	VGWTAGLSSFAA	IPFAQISIF	YRLNGVIGITQVLS	ENQKLIANKFNQALGAMQTGFTTNEAFRKVQDAVNNAQALSKLASELSNTFGAISASIGDIIQRLDVLQDAQIDRLINGRLTT		1077												
	*	:	..*	:	..*	*	*	*	*	*	*	*	*	*	*	*	*	*
HKU1	LNAVVSQQLSDITLIKAGASRAIEK	VNECVKSQSPRINFCGNGNHLISLVQNA	PYGLLFIHFSYKPTSEKTVLVS	PGCLSGDR--	GIAPKQGYFIQK-----	NDSWMFTGSSYYYPEPI	1197											
SARS	LQTYVTQQLIRAAEIRASANLAAT	KMSECVLQSKRVD	FCGKGHYHLSFPQAAPHGVVFLHVTYVPSQERNFTTAPACHEGKA--	YFPREGVVFVN-----	GTSWFTQNRNFSQII	1097												
MERS	LNAFVAQQLVRSESAALSAQLAKD	KVNECVKAQSKRSGFCGQGT	HIVSVVNA	PNGLYFMHVGYPSNHIEVVSAYGLCDAANPTNCIAPVNGYFIKTN	NRTRIVDEWSYTGSSFYSEPEPI	1197												
	*:::.*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
			HR2	TM														
HKU1	SDKNVFMNSC	SVNFTKAPFIYL--	NNSIPNLSDFEAE	LSLWFKNHTSIAPNLT	FNSHINATFLDLYYEMNVIQESIKSLNSSFINLKEIGTYEMV	KWPWYIWLILVILFIIFLMILFF	1315											
SARS	TTDNTFVSGNCDV	VIIGINNIVYDPLQ--	PELDSFKEELDKYFNKHTSPD	VDLGDISGINATFLDLYYEMNVIQESIKSLNSSFINLKEIGTYEMV	KWPWYIWLILVILFIIFLMILFF		1215											
MERS	TSLNTRYVAPQVT-YQ	NISTNLPPLLG	NSTGIDFQDELDEFFKNVSTSI	PNFSGSLTQINTLLDLTYEMLSLQOVV	KALNESYIDLKELGNYTYNKKWPWYIWLGLFIAGLVALALCVFF		1316											
	:	*	:	..*	::	:	:	:	:	:	:	:	:	:	:	:	:	:
HKU1	ICCCTGCGSACFSK--	CHNCCDEYGGHNDEVIK-ASHDD--																
SARS	LCCMTSCCS	SLKGACSCGCKFD-EDDSE	PVLKGVLKGLHYT															
MERS	ILCCTGCGTNC	MGLKCRNCCDRYEEYDLEPHKVH	VH----															
	:	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

**Extended Data Figure 6 | Sequence alignment of human betacoronavirus S proteins.** Sequence alignment of S proteins from HKU1, SARS-CoV and MERS-CoV using Clustal Omega<sup>52</sup>. Protein features described in the text are indicated: N-terminal domain (NTD),

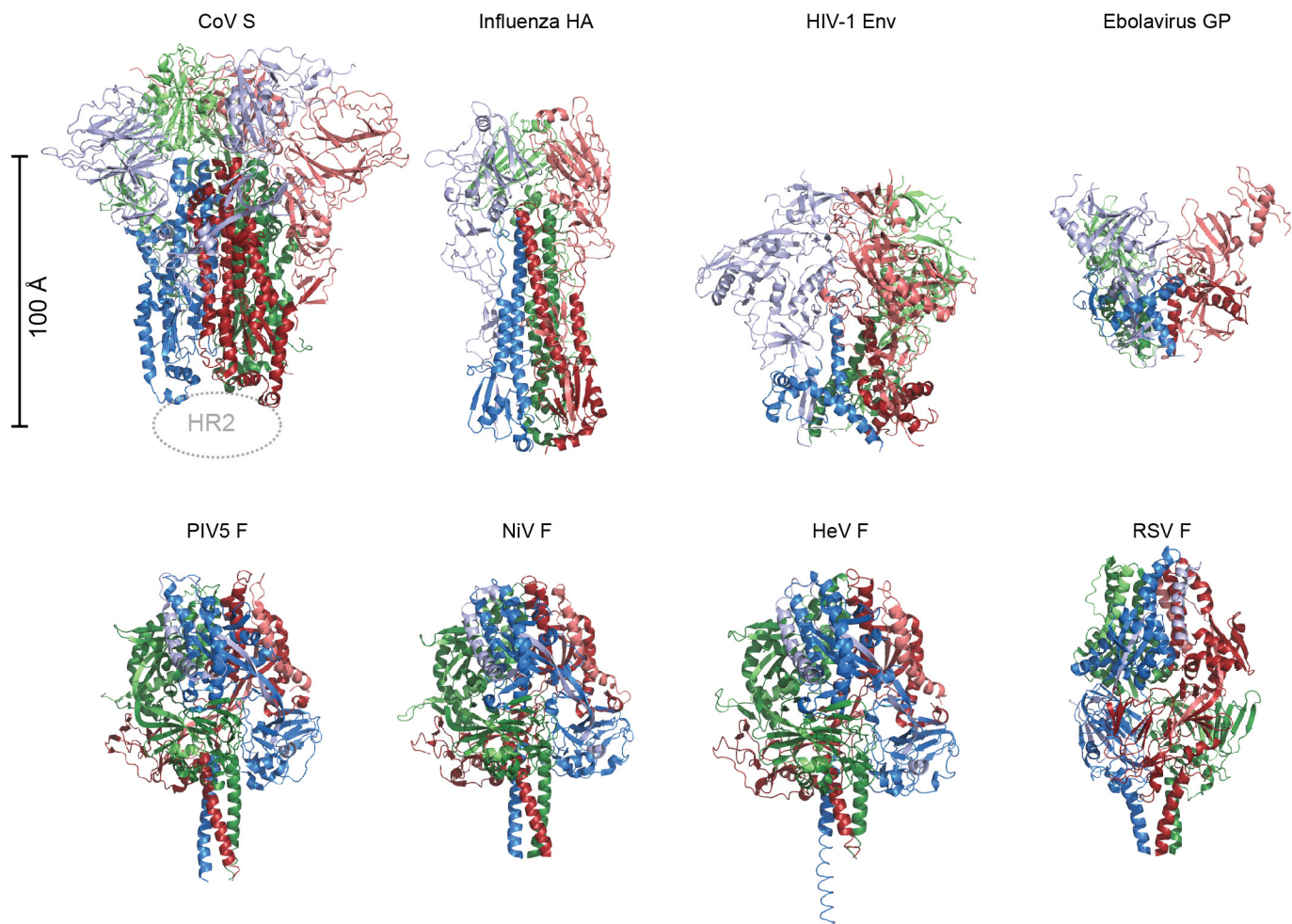
C-terminal domain (CTD) which contains the large variable loop, the S1/S2 and S2' cleavage sites, fusion peptide (FP), heptad repeats 1 and 2 (HR1, HR2) and transmembrane helix (TM).





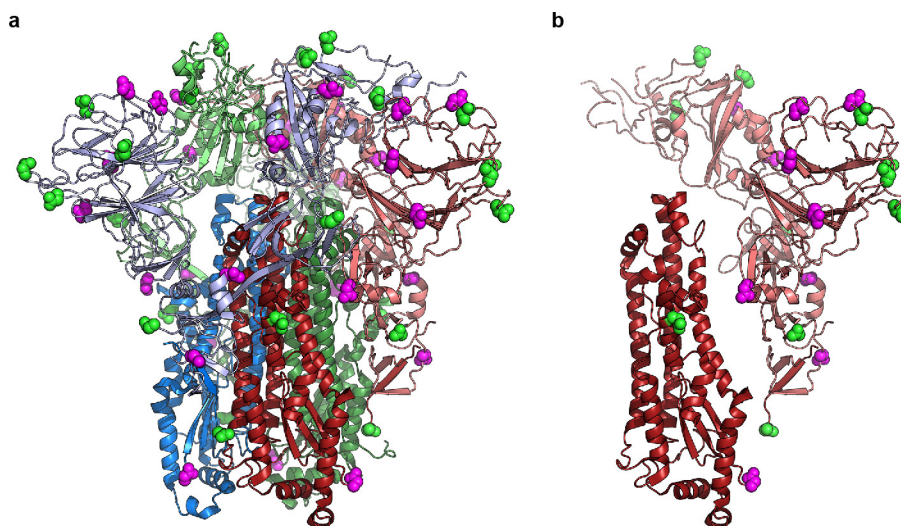
**Extended Data Figure 7 | S1 sits atop an adjacent protomer's S2.** **a**, The HKU1 S1 subunits are rotated about the trimeric threefold axis relative to their corresponding S2 subunits such that the S1 CTD from one protomer caps the S2 central helix from an adjacent protomer (CTD<sub>1</sub>, blue, caps S2<sub>2</sub>, red). The third protomer of the trimer has been omitted for clarity. **b**, HKU1 S1 CTD (blue) uses a short helix to cap the central helix and HR1 (red). **c**, The influenza haemagglutinin HA2 central helix (red) is also capped by a helix in HA1 (blue)<sup>14,28</sup>. **d**, The S2 N-terminal  $\beta$ -strand is connected to the remainder of the S2 subunit via a loop and an  $\alpha$ -helix

(dotted lines). These regions of the EM density are of insufficient quality to confidently build this protein region but enable interpretation of connectivity. **e**, In the pre-fusion HKU1 S protein, the tops of the central S2 helices (blue, red, green) are splayed outwards from the threefold axis and capped by the S1 CTDs (white). The S1 NTD, SD-1 and SD-2 have been omitted for clarity. **f**, In the post-fusion six-helix-bundle structure of SARS S<sup>22</sup>, the corresponding helical regions from (e) form a well-packed three-helix bundle.



**Extended Data Figure 8 | Class I viral fusion proteins.** All class I fusion proteins require proteolytic cleavage adjacent to the fusion peptide or loop, and the metastable pre-fusion state is triggered by a series of events that involve pH change or receptor binding. The post-fusion conformations all contain anti-parallel six-helix bundles composed of the HR1 and HR2 from the membrane-proximal subunit. However, there is a great diversity in pre-fusion conformations as shown here. Members of this class that also participate in receptor binding<sup>14–16,28,53</sup> (top row), including

S glycoproteins of coronaviruses, are organized such that their receptor binding subunits sit atop the fusion machinery, and need to be shed in order for membrane fusion to proceed. Paramyxovirus F proteins<sup>54–57</sup> (bottom row) have a different architecture than the capped fusion proteins on the top row. The F proteins all have disulfide bonds between the membrane proximal and membrane distal subunits, and the two subunits remain interconnected throughout the rearrangement process.



**Extended Data Figure 9 | HKU1 S glycosylation.** **a**, Sites of *N*-linked glycosylation on the HKU1 S trimer and **b**, a single monomer. Of the 30 potential *N*-linked glycosylation sites in a single protomer, the asparagine residues are observed for 21 sites and of these a small portion

of density in the EM map is observed for 10 sites corresponding to the EndoH-trimmed sugars. Asparagines where glycan density is observed are shown as magenta spheres. Asparagines lacking glycan density are shown in green.



Extended Data Table 1 | CryoEM data collection, processing and refinement metrics

**Data collection/processing**

Microscope	Titan Krios
Voltage (keV)	300
Defocus range ( $\mu\text{m}$ )	1.0 to 3.5
Movies	1,049
Frames per movie	50
Exposure time per frame (ms)	200
Magnification	22,500x
Dose rate ( $\text{e}^-/\text{pixel}/\text{s}$ )	10
Total dose per movie ( $\text{e}^-/\text{\AA}^2$ )	57
Particles	31,435
Map Resolution ( $\text{\AA}$ )	4.04

**Model Refinement**

Chimera CC <sup>44</sup>	0.87
EMRinger Score <sup>48</sup>	2.7
MolProbity <sup>49</sup>	1.6
Clashscore <sup>49</sup>	3.0
Ramachandran (%) <sup>49</sup>	
Favored	92.1
Allowed	7.0
Outliers	0.9

CC = cross correlation

# Crystal structure of eukaryotic translation initiation factor 2B

Kazuhiro Kashiwagi<sup>1,2,3</sup>, Mari Takahashi<sup>3</sup>, Madoka Nishimoto<sup>3</sup>, Takuya B. Hiyama<sup>1,2</sup>, Toshiaki Higo<sup>2,3</sup>, Takashi Umehara<sup>2,3</sup>, Kensaku Sakamoto<sup>2,3</sup>, Takuhiro Ito<sup>1,2,3</sup> & Shigeyuki Yokoyama<sup>1,2,4</sup>

Eukaryotic cells restrict protein synthesis under various stress conditions, by inhibiting the eukaryotic translation initiation factor 2B (eIF2B)<sup>1,2</sup>. eIF2B is the guanine nucleotide exchange factor for eIF2, a heterotrimeric G protein consisting of  $\alpha$ -,  $\beta$ - and  $\gamma$ -subunits. eIF2B exchanges GDP for GTP on the  $\gamma$ -subunit of eIF2 (eIF2 $\gamma$ ), and is inhibited by stress-induced phosphorylation of eIF2 $\alpha$ . eIF2B is a heterodecameric complex of two copies each of the  $\alpha$ -,  $\beta$ -,  $\gamma$ -,  $\delta$ - and  $\epsilon$ -subunits<sup>3</sup>; its  $\alpha$ -,  $\beta$ - and  $\delta$ -subunits constitute the regulatory subcomplex<sup>4</sup>, while the  $\gamma$ - and  $\epsilon$ -subunits form the catalytic subcomplex<sup>5</sup>. The three-dimensional structure of the entire eIF2B complex has not been determined. Here we present the crystal structure of *Schizosaccharomyces pombe* eIF2B with an unprecedented subunit arrangement, in which the  $\alpha_2\beta_2\delta_2$  hexameric regulatory subcomplex binds two  $\gamma\epsilon$  dimeric catalytic subcomplexes on its opposite sides. A structure-based *in vitro* analysis by a surface-scanning site-directed photo-cross-linking method identified the eIF2 $\alpha$ -binding and eIF2 $\gamma$ -binding interfaces, located far apart on the regulatory and catalytic subcomplexes, respectively. The eIF2 $\gamma$ -binding interface is located close to the conserved 'NF motif', which is important for nucleotide exchange. A structural model was constructed for the complex of eIF2B with phosphorylated eIF2 $\alpha$ , which binds to eIF2B more strongly than the unphosphorylated form. These results indicate that the eIF2 $\alpha$  phosphorylation generates the 'nonproductive' eIF2-eIF2B complex<sup>5</sup>, which prevents nucleotide exchange on eIF2 $\gamma$ , and thus provide a structural framework for the eIF2B-mediated mechanism of stress-induced translational control.

In eukaryotic translation initiation, eIF2 in the GTP-bound form delivers an initiator methionyl-tRNA (Met-tRNA<sup>Met</sup>) to the ribosome, and then dissociates in the GDP-bound form<sup>2</sup>. For the next round of translation initiation, eIF2B catalyses the exchange of the eIF2 $\gamma$ -bound GDP for GTP. This guanine nucleotide exchange activity requires the HEAT domain at the carboxy (C) terminus and the NF motif in the amino (N)-terminal region of the  $\epsilon$ -subunit, as well as the formation of the catalytic  $\gamma\epsilon$  subcomplex<sup>5-7</sup>. The NF motif consists of consecutive Asn-Phe residues conserved in eIF2B $\epsilon$  (Asn237 and Phe238 in *S. pombe* eIF2B $\epsilon$ ), and their mutations reduce the nucleotide exchange activity drastically, down to the level of a HEAT-domain fragment<sup>7</sup>. Under stressful conditions, the phosphorylation of eIF2 $\alpha$  at Ser51 induces its stronger binding to the regulatory  $\alpha\beta\delta$  subcomplex of eIF2B, which results in the inhibition of eIF2B<sup>4,8,9</sup>. This inhibition limits the supply of Met-tRNA<sup>Met</sup> to the ribosome, leading to global translational repression and de-repression of the translation of stress-induced mRNAs<sup>1,2,8</sup>. However, the mechanisms of the catalysis and the inhibition and their mutual relationship have remained elusive, and information about the overall structure of eIF2B has long been awaited. Importantly, a variety of mutations of the human eIF2B subunits are related to the neurodegenerative disease leukoencephalopathy with vanishing white matter (VWM) or childhood ataxia with central

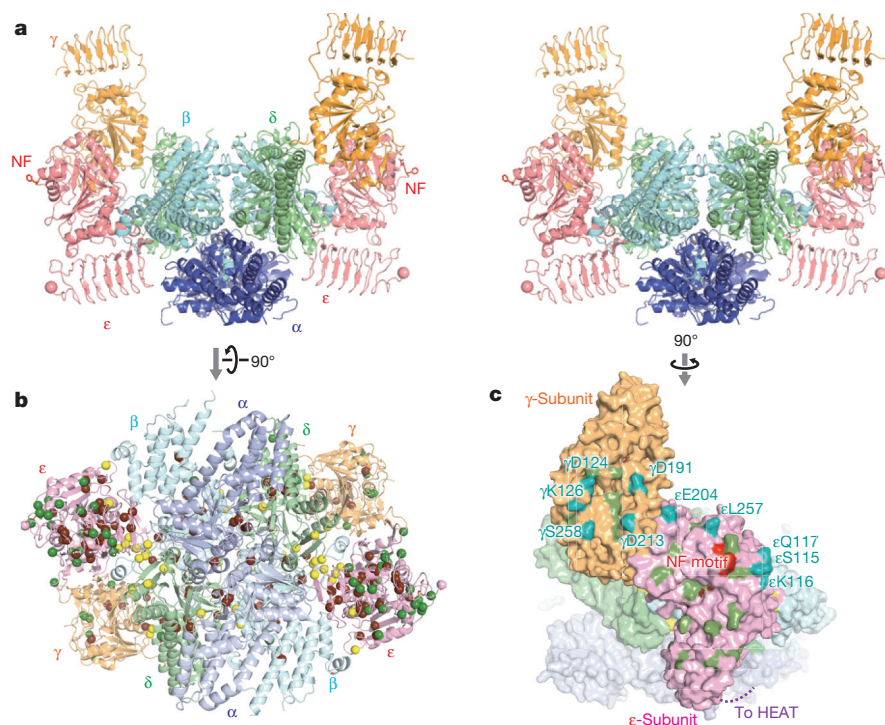
nervous system hypomyelination (CACH)<sup>10</sup>. In patients with this disease, white matter lesions and neurological disorders severely deteriorate during recovery after exposure to stresses, and the eIF2B activities are generally lower than normal<sup>11</sup>. Thus, the structure of eIF2B would also provide information underlying the pathogenesis of VWM/CACH.

We prepared *S. pombe* eIF2B, by co-expressing all subunits in *Escherichia coli* (Kashiwagi *et al.*, submitted). We examined the nucleotide exchange activity against *Komagataella pastoris* (*Pichia pastoris*) eIF2, which shares high sequence identity with *S. pombe* eIF2 ( $\alpha$ : 66%;  $\beta$ : 47%;  $\gamma$ : 76%). This recombinant eIF2B molecule exhibited the nucleotide exchange activity on *K. pastoris* eIF2, and was inhibited by the phosphorylated eIF2 protein (eIF2( $\alpha$ P)) (Extended Data Fig. 1a). The eIF2B molecule showed stronger binding to both the trimeric eIF2( $\alpha$ P) complex and the phosphorylated eIF2 $\alpha$  subunit (P-eIF2 $\alpha$ ) (Extended Data Fig. 1b, c). Therefore, the recombinant eIF2B molecule displayed the characteristic biochemical properties of the natural eIF2B.

We determined the crystal structure of eIF2B at 3.0-Å resolution, and assigned almost all regions except for the  $\epsilon$ -subunit HEAT domains (Fig. 1a, Extended Data Fig. 2a, b and Extended Data Table 1). The crystal structure revealed an unprecedented arrangement of the subunits: the hexameric regulatory subcomplex resides at the centre, with the two heterodimeric catalytic subcomplexes bound on opposite sides. The assembly of the subcomplexes is primarily mediated by the  $\beta$ - $\epsilon$  and  $\delta$ - $\gamma$  interactions (Fig. 1a and Extended Data Fig. 2a, b). This decameric structure is consistent with many of the numerous previously reported results of subunit interactions<sup>12-16</sup> (Extended Data Fig. 2). Mapping of the residues corresponding to the missense mutations causing VWM disease (Fig. 1b and Supplementary Table 1) revealed that many mutations are located within or around various subunit interfaces (yellow in Fig. 1b and Extended Data Fig. 3a-c). These interface mutations may cause substantial effects on the structural and biochemical properties of eIF2B.

Two key motifs for the nucleotide exchange activity, the HEAT domain and the NF motif of the  $\epsilon$ -subunit, both reside in the 'distal' region of the structure, and several missense VWM mutations are mapped near the NF motif (Fig. 1c and Extended Data Fig. 3d). Therefore, nucleotide exchange on eIF2 $\gamma$  occurs on the distal face of eIF2B. To examine how eIF2 $\gamma$  binds to this face, we performed surface-scanning photo-cross-linking experiments (see Methods). Thirty-two variants of eIF2B, labelled with *p*-benzoyl-L-phenylalanine (pBpa), were ultraviolet-irradiated in the presence of *K. pastoris* eIF2 (Extended Data Fig. 4a, b). Photo-cross-linking was detected between eIF2 $\gamma$  and the distal face of eIF2B, when pBpa was incorporated at one of the ten sites distributed over a large area, extending from eIF2B $\epsilon$  to eIF2B $\gamma$  (Fig. 1c). The phosphorylation of eIF2 retarded cross-linking at the sites near the NF motif (Gln117(2B $\epsilon$ ) and Leu257(2B $\epsilon$ )), demonstrating that the interaction of the  $\gamma$ -subunit of eIF2( $\alpha$ P) with the NF motif-surrounding region of eIF2B $\epsilon$  is much less efficient than that of the unphosphorylated form (Extended Data Fig. 4b, c). This observation

<sup>1</sup>Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan. <sup>2</sup>RIKEN Systems and Structural Biology Center, Tsurumi-ku, Yokohama 230-0045, Japan. <sup>3</sup>RIKEN Center for Life Science Technologies, Tsurumi-ku, Yokohama 230-0045, Japan. <sup>4</sup>RIKEN Structural Biology Laboratory, Tsurumi-ku, Yokohama 230-0045, Japan.



**Figure 1 | Overall structure of *S. pombe* eIF2B and mapping of human VWM mutations and eIF2 $\gamma$  cross-links. a**, The crystal structure of *S. pombe* eIF2B (wall-eyed stereo view). The  $\alpha$ -,  $\beta$ -,  $\gamma$ -,  $\delta$ - and  $\epsilon$ -subunits are coloured blue, cyan, orange, green and pink, respectively. The NF motifs in the  $\epsilon$ -subunits are shown by red sticks. The visible C-terminal C $\alpha$  atoms of the  $\epsilon$ -subunits are shown by spheres, and the HEAT domains, whose electron densities were not observed in our structure, extend outwards. **b**, Mapping of the *S. pombe* eIF2B residues corresponding to

VWM-causing missense mutations of human eIF2B (Supplementary Table 1) as spheres on the *S. pombe* eIF2B structure. Their environments are colour-coded on the spheres (green, solvent exposed; yellow, subunit interface; brown, structural core). **c**, Mapping of the *S. pombe* eIF2B residues corresponding to the solvent-exposed VWM-causing missense mutations of human eIF2B (green) and to the pBpa cross-links with *K. pastoris* eIF2 $\gamma$ , through the interaction between eIF2B and eIF2 (teal), on the surface model. The NF motif is coloured red.

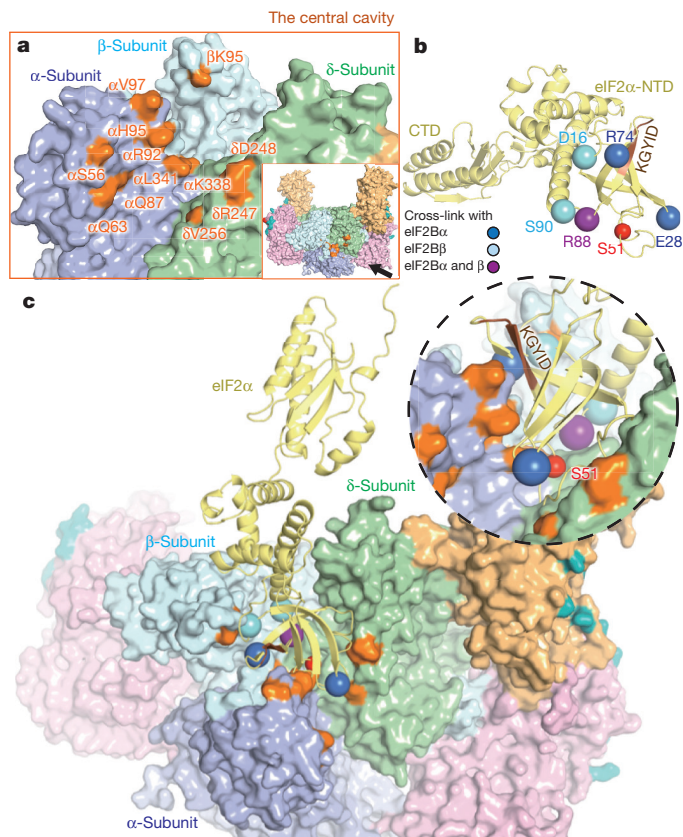
may indicate why eIF2( $\alpha$ P) is a poor substrate for nucleotide exchange by eIF2B. Meanwhile, cross-linking at the sites distant from the NF motif (Glu204(2B $\epsilon$ ) and Ser258(2B $\gamma$ )) was hardly affected by the eIF2 phosphorylation (Extended Data Fig. 4b, c). The different effects of the eIF2 phosphorylation may reflect the dual functions of the catalytic subcomplex on eIF2 (ref. 17). The catalytic subcomplex also catalyses the displacement of eIF5, which detaches from the ribosome together with eIF2 and inhibits the dissociation of eIF2 $\gamma$ -bound GDP (ref. 18). This activity is insensitive to phosphorylation and impaired by eIF2B $\gamma$  mutations<sup>17</sup>. Therefore, the observed phosphorylation-insensitive cross-linking at the distant sites is likely to have trapped the interaction for this step. Furthermore, several VWM mutations are mapped on the same surface area of the eIF2B $\gamma$  subunit as the phosphorylation-insensitive cross-linking (Fig. 1c and Extended Data Fig. 3d), suggesting that defects in the eIF5 displacement are relevant to VWM disease in these cases.

We next investigated the interaction between the eIF2B regulatory subcomplex and P-eIF2 $\alpha$ . pBpa was incorporated at 80 sites in the regulatory  $\alpha$ -,  $\beta$ - and  $\delta$ -subunits, and 13 of them were cross-linked with *S. pombe* P-eIF2 $\alpha$ . The cross-linked sites are distributed in all three regulatory subunits, and are clustered within the cavity-like regions of the eIF2B regulatory subcomplex (Fig. 2a and Extended Data Fig. 5a, b). The cavities are formed around the centre of one set of the  $\alpha$ -,  $\beta$ - and  $\delta$ -subunits ('central cavity') on the top/bottom ends of the hexameric regulatory subcomplex (Fig. 1a). Thus, eIF2B possesses two P-eIF2 $\alpha$ -binding sites, in agreement with the ITC results (Extended Data Fig. 1c). This cavity bears some residues with mutations that have been isolated as Gcn<sup>-</sup> (general control non-depressible) mutations<sup>19,20</sup>, which prevent cells from inducing translational control upon eIF2 phosphorylation, and their positions overlap with the eIF2 $\alpha$ -cross-linking sites (Extended Data Fig. 6a–c). In addition, the majority of the Gcn<sup>-</sup> mutation positions are located at the interfaces between the

regulatory dimers (Extended Data Fig. 6d, e), demonstrating that the correct assembly of the regulatory subunits is requisite for the strong binding with P-eIF2 $\alpha$  and the induction of translational control. In contrast, fewer missense VWM mutations were identified in the central cavity (Extended Data Fig. 3f). Therefore, it seems that VWM disease occurs in most cases by mechanisms unrelated to the inhibition of the GEF activity of eIF2B by the eIF2 $\alpha$  phosphorylation. Intriguingly, a similar cross-linking pattern was observed without eIF2 $\alpha$  phosphorylation (Extended Data Fig. 5c), although the binding was weaker. The difference was that additional cross-links occurred at Arg84(2B $\beta$ ) and Gln91(2B $\beta$ ) when eIF2 $\alpha$  was not phosphorylated, and both of these residues exist in the interior of the cavity (Extended Data Fig. 5d). This suggests that the central cavity can accommodate both phosphorylated and unphosphorylated eIF2 $\alpha$  in similar manners, but the interaction with eIF2 $\alpha$  is somewhat delocalized in the absence of phosphorylation.

To delineate how eIF2 $\alpha$  is accommodated in the central cavity, we introduced pBpa at 27 sites in the N-terminal domain of *S. pombe* eIF2 $\alpha$  (eIF2 $\alpha$ -NTD) and examined the cross-linking with the eIF2B regulatory subunits. We successfully detected the cross-linking of eIF2 $\alpha$  with eIF2B $\alpha$  and eIF2B $\beta$  (Extended Data Fig. 7a, b). Mapping of the cross-linked sites onto the human eIF2 $\alpha$  structure<sup>21</sup> revealed that the tip of eIF2 $\alpha$ -NTD is located close to eIF2B $\alpha$  and eIF2B $\beta$  (Fig. 2b). On the basis of these complementary experiments, the structures of eIF2 $\alpha$  and eIF2B were docked in accordance with the cross-linking results (Fig. 2c), which indicated that eIF2 $\alpha$ -NTD is buried in the central cavity along with the notch between the  $\beta$ - and  $\delta$ -subunits of eIF2B. The eIF2 $\alpha$  KGYID sequence (residues 79–83), which is important for the strong binding to eIF2B<sup>9,22</sup>, closely faces eIF2B $\alpha$  (Fig. 2c, inset). The residues Ser48, Leu84 and Val89 of eIF2 $\alpha$ , whose mutations suppress the inhibitory effect of the eIF2 $\alpha$  phosphorylation<sup>23</sup>, are located at the bottom of the cavity. Therefore, our docking model seems to be reasonable. In the model, the phosphorylated residue Ser51(2 $\alpha$ ) is also located at the





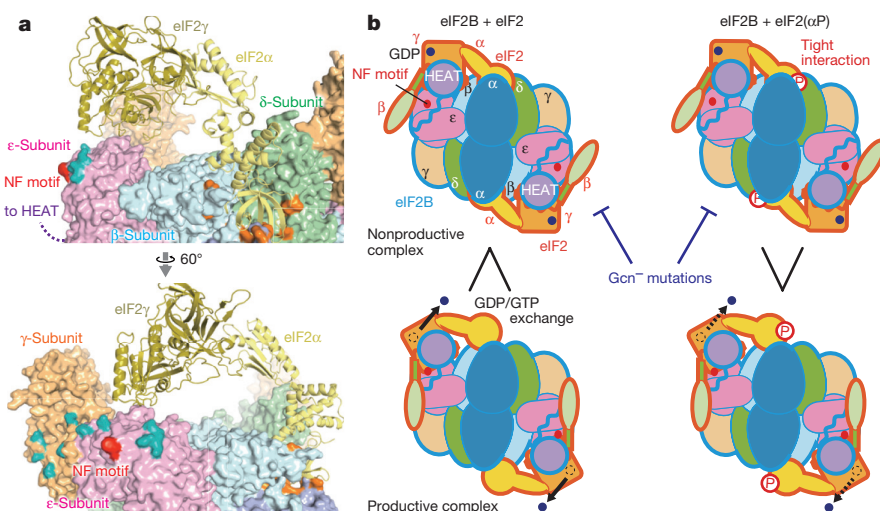
**Figure 2 | Photo-cross-linking between eIF2 $\alpha$  and the central cavity of the eIF2B regulatory subunit.** **a**, Mapping of the cross-linked (orange) eIF2B sites with *S. pombe* eIF2 $\alpha$  on the crystal structure (surface model) (see Extended Data Fig. 5a). Asn100(2B $\beta$ ) is disordered in this structure. Inset: the same view as in Fig. 1a (surface model). **b**, Mapping of the eIF2 $\alpha$  sites cross-linked with eIF2B $\alpha$  and eIF2B $\beta$  on the human eIF2 $\alpha$  structure<sup>21</sup> (ribbon model; see Extended Data Fig. 7a, b). The sites cross-linked with eIF2B $\alpha$ , eIF2B $\beta$  and both eIF2B $\alpha$  and eIF2B $\beta$  are represented as blue, cyan and purple spheres, respectively. The phosphorylated residue Ser51(2 $\alpha$ ) is represented as the red sphere. The KGYID sequence is coloured brown. **c**, Docking of the eIF2B and eIF2 $\alpha$  structures, based on our photo-cross-linking experiments. Inset: the same view as in **a**.

bottom of the cavity (Fig. 2c, inset). The cavity has no positively charged patch suitable for the accommodation of a phosphate group (Extended Data Fig. 7c, d). Therefore, the mechanism underlying the enhanced affinity for P-eIF2 $\alpha$  is still unclear.

Considering our finding that the cross-linking at the sites around the NF motif is impaired by the eIF2 phosphorylation, as described above, eIF2 $\gamma$  seems to minimally contact the NF motif when P-eIF2 $\alpha$ -NTD binds to the central cavity. On the basis of the docking model of P-eIF2 $\alpha$ -NTD captured by the central cavity, we further docked the trimeric aIF2, the archaeal homologue of eIF2 (ref. 24), on the eIF2B structure. This docking model indicates that it is difficult for the aIF2/eIF2 captured by the central cavity to simultaneously interact with the NF motif (Fig. 3a). Therefore, the central-cavity-captured state is the ‘nonproductive’ state, which is distinct from the ‘productive’ state for efficient nucleotide exchange (Fig. 3b, left)<sup>5</sup>. When eIF2 $\alpha$  is not phosphorylated, eIF2 $\gamma$ -bound GDP is exchanged with GTP by the HEAT domain and the NF motif, because the nonproductive state is not stable (Fig. 3b, left). The stress-induced eIF2 $\alpha$  phosphorylation stabilizes the nonproductive state by the stronger interaction in the central cavity, thereby impeding the nucleotide exchange on the bound eIF2 and the entry of another eIF2 (Fig. 3b, right). According to this model, the Gcn<sup>−</sup> mutations, which alleviate translational control upon eIF2 phosphorylation, should destabilize the strong interaction in the central cavity. We selected two Gcn<sup>−</sup> mutations in the central cavity, Glu57Lys(2B $\alpha$ ) and Asp248Lys(2B $\delta$ ) (Glu44Lys(2B $\alpha$ ) and Glu377Lys(2B $\delta$ ) in *Saccharomyces cerevisiae*) (Extended Data Fig. 8a). These eIF2B mutations abrogated its strong interactions with eIF2( $\alpha$ P) and P-eIF2 $\alpha$  (Extended Data Fig. 8b, c). The GDP exchange assay also revealed the alleviation of the inhibition by eIF2( $\alpha$ P) (Extended Data Fig. 8d). Therefore, these Gcn<sup>−</sup> mutations diminish the strong interaction in the central cavity, thus restoring efficient nucleotide exchange (Fig. 3b).

These analyses additionally revealed that the Gcn<sup>−</sup> mutations also affect the interaction with unphosphorylated eIF2 $\alpha$ , and enhance the nucleotide exchange activity of eIF2B (Extended Data Fig. 8c, e). The nucleotide exchange activity of eIF2B is also enhanced by a small molecule called integrated stress response inhibitor (ISRIB)<sup>25,26</sup>, which was identified as a compound that blocks the stress response<sup>27</sup>. The ISRIB-resistant mutations<sup>26</sup> are located near the pseudo-twofold rotational axis of eIF2B (Extended Data Fig. 9). Therefore, our structure is consistent with the proposed models of action of ISRIB, in which this symmetric molecule stabilizes the eIF2B decamer<sup>25,26</sup>.

The current study of eIF2B provides the structural basis not only for an overall understanding of its function in translational control, but also of the mechanisms of VWM. We have focused so far on the structural integrity and the catalytic activities of nucleotide exchange and eIF5 displacement, as the causes of this disease. However, VWM disease can reportedly result from alterations in other, as yet undescribed, eIF2B functions<sup>28,29</sup>. Mapping of missense VWM mutations on the structure actually revealed the existence of some exposed mutations



**Figure 3 | Model of the eIF2-eIF2B interactions.**

**a**, Docking of the eIF2 and eIF2B structures, by fixing eIF2 $\alpha$ -NTD in the central cavity of eIF2B. eIF2 $\alpha$ -NTD is placed in the same manner as in Fig. 2c. The structures of aIF2 $\alpha$ -CTD and aIF2 $\gamma$  from PDB accession number 2QMU<sup>24</sup>, the archaeal homologues of the corresponding segments of eIF2, are positioned so they extend towards the NF motif (red) from the C terminus of eIF2 $\alpha$ -NTD. **b**, Schematic representations of the eIF2-eIF2B interactions proposed in the present study.

that may not be explained by defects in the known functions (for example,  $\gamma$ Arg312Gln and  $\gamma$ Ile346Thr in Extended Data Fig. 3b and the mutations in Extended Data Fig. 3e). The present decamer structure is expected to contribute to the elucidation of the currently undescribed mechanisms of this disease.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 9 October 2015; accepted 12 January 2016.**

**Published online 22 February 2016.**

- Pavitt, G. D. eIF2B, a mediator of general and gene-specific translational control. *Biochem. Soc. Trans.* **33**, 1487–1492 (2005).
- Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Rev. Mol. Cell Biol.* **11**, 113–127 (2010).
- Kuhle, B., Eulig, N. K. & Ficner, R. Architecture of the eIF2B regulatory subcomplex and its implications for the regulation of guanine nucleotide exchange on eIF2. *Nucleic Acids Res.* **43**, 9994–10014 (2015).
- Yang, W. & Hinnebusch, A. G. Identification of a regulatory subcomplex in the guanine nucleotide exchange factor eIF2B that mediates inhibition by phosphorylated eIF2. *Mol. Cell. Biol.* **16**, 6603–6616 (1996).
- Pavitt, G. D., Ramaiah, K. V. A., Kimball, S. R. & Hinnebusch, A. G. eIF2 independently binds two distinct eIF2B subcomplexes that catalyze and regulate guanine-nucleotide exchange. *Genes Dev.* **12**, 514–526 (1998).
- Gomez, E., Mohammad, S. S. & Pavitt, G. D. Characterization of the minimal catalytic domain within eIF2B: the guanine-nucleotide exchange factor for translation initiation. *EMBO J.* **21**, 5292–5301 (2002).
- Gomez, E. & Pavitt, G. D. Identification of domains and residues within the  $\epsilon$  subunit of eukaryotic translation initiation factor 2B (eIF2B $\epsilon$ ) required for guanine nucleotide exchange reveals a novel activation function promoted by eIF2B complex formation. *Mol. Cell. Biol.* **20**, 3965–3976 (2000).
- Wek, R. C., Jiang, H. Y. & Anthony, T. G. Coping with stress: eIF2 kinases and translational control. *Biochem. Soc. Trans.* **34**, 7–11 (2006).
- Krishnamoorthy, T., Pavitt, G. D., Zhang, F., Dever, T. E. & Hinnebusch, A. G. Tight binding of the phosphorylated  $\alpha$  subunit of initiation factor 2 (eIF2 $\alpha$ ) to the regulatory subunits of guanine nucleotide exchange factor eIF2B is required for inhibition of translation initiation. *Mol. Cell. Biol.* **21**, 5018–5030 (2001).
- Pavitt, G. D. & Proud, C. G. Protein synthesis and its control in neuronal cells with a focus on vanishing white matter disease. *Biochem. Soc. Trans.* **37**, 1298–1310 (2009).
- Fogli, A. & Boespflug-Tanguy, O. The large spectrum of eIF2B-related diseases. *Biochem. Soc. Trans.* **34**, 22–29 (2006).
- Wang, X., Wortham, N. C., Liu, R. & Proud, C. G. Identification of residues that underpin interactions within the eukaryotic initiation factor (eIF2) 2B complex. *J. Biol. Chem.* **287**, 8263–8274 (2012).
- Reid, P. J., Mohammad-Qureshi, S. S. & Pavitt, G. D. Identification of intersubunit domain interactions within eukaryotic initiation factor (eIF) 2B, the nucleotide exchange factor for translation initiation. *J. Biol. Chem.* **287**, 8275–8285 (2012).
- Wortham, N. C., Martinez, M., Gordiyenko, Y., Robinson, C. V. & Proud, C. G. Analysis of the subunit organization of the eIF2B complex reveals new insights into its structure and regulation. *FASEB J.* **28**, 2225–2237 (2014).
- Bogorad, A. M. *et al.* Insights into the architecture of the eIF2B $\alpha$ / $\beta$ / $\delta$  regulatory subcomplex. *Biochemistry* **53**, 3432–3445 (2014).
- Gordiyenko, Y. *et al.* eIF2B is a decameric guanine nucleotide exchange factor with a  $\gamma$  $\epsilon$  $\delta$  tetrameric core. *Nature Commun.* **5**, 3902 (2014).
- Jennings, M. D., Zhou, Y., Mohammad-Qureshi, S. S., Bennett, D. & Pavitt, G. D. eIF2B promotes eIF5 dissociation from eIF2•GDP to facilitate guanine nucleotide exchange for translation initiation. *Genes Dev.* **27**, 2696–2707 (2013).
- Jennings, M. D. & Pavitt, G. D. eIF5 has GDI activity necessary for translational control by eIF2 phosphorylation. *Nature* **465**, 378–381 (2010).
- Vazquez de Aldana, C. R. & Hinnebusch, A. G. Mutations in the GCD7 subunit of yeast guanine nucleotide exchange factor eIF-2B overcome the inhibitory effects of phosphorylated eIF-2 on translation initiation. *Mol. Cell. Biol.* **14**, 3208–3222 (1994).
- Pavitt, G. D., Yang, W. & Hinnebusch, A. G. Homologous segments in three subunits of the guanine nucleotide exchange factor eIF2B mediate translational regulation by phosphorylation of eIF2. *Mol. Cell. Biol.* **17**, 1298–1313 (1997).
- Ito, T., Marintchev, A. & Wagner, G. Solution structure of human initiation factor eIF2 $\alpha$  reveals homology to the elongation factor eEF1B. *Structure* **12**, 1693–1704 (2004).
- Dey, M. *et al.* PKR and GCN2 kinases and guanine nucleotide exchange factor eukaryotic translation initiation factor 2B (eIF2B) recognize overlapping surfaces on eIF2 $\alpha$ . *Mol. Cell. Biol.* **25**, 3063–3075 (2005).
- Vazquez de Aldana, C. R., Dever, T. E. & Hinnebusch, A. G. Mutations in the  $\alpha$  subunit of eukaryotic translation initiation factor 2 (eIF-2 $\alpha$ ) that overcome the inhibitory effect of eIF-2 $\alpha$  phosphorylation on translation initiation. *Proc. Natl Acad. Sci. USA* **90**, 7215–7219 (1993).
- Yatime, L., Mechulam, Y., Blanquet, S. & Schmitt, E. Structure of an archaeal heterotrimeric initiation factor 2 reveals a nucleotide state between the GTP and the GDP states. *Proc. Natl Acad. Sci. USA* **104**, 18445–18450 (2007).
- Sidrauski, C. *et al.* Pharmacological dimerization and activation of the exchange factor eIF2B antagonizes the integrated stress response. *eLife* **4**, e07314 (2015).
- Sekine, Y. *et al.* Stress responses. Mutations in a translation initiation factor identify the target of a memory-enhancing compound. *Science* **348**, 1027–1030 (2015).
- Sidrauski, C. *et al.* Pharmacological brake-release of mRNA translation enhances cognitive memory. *eLife* **2**, e00498 (2013).
- Liu, R. *et al.* Severity of vanishing white matter disease does not correlate with deficits in eIF2B activity or the integrity of eIF2B complexes. *Hum. Mutat.* **32**, 1036–1045 (2011).
- Wortham, N. C. & Proud, C. G. Biochemical effects of mutations in the gene encoding the alpha subunit of eukaryotic initiation factor (eIF) 2B associated with vanishing white matter disease. *BMC Med. Genet.* **16**, 64 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the staff of the beamline BL41XU at SPring-8 for their support. This work was performed with the approval of the Japan Synchrotron Radiation Research Institute (proposals 2012A1335 and 2012B1572). This work was supported by JSPS KAKENHI grants 23687013 and 25121737 (to T.I.), the Targeted Proteins Research Program (TPRP) and the Platform for Drug Discovery, Informatics and Structural Life Science, from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (to S.Y.), and a research program of the UT-RIKEN Cooperation Laboratory of Structural Biology (to S.Y.).

**Author Contributions** K.K. and T.I. designed and performed experiments, and wrote the manuscript. M.T., M.N. and T.B.H. purified the samples. T.H. and T.U. established the *K. pastoris* strain. K.S. provided the materials and technical advice for the surface-scanning photo-cross-linking experiments. S.Y. designed experiments and wrote the manuscript.

**Author Information** The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession number 5B04. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.Y. ([yokoyama@riken.jp](mailto:yokoyama@riken.jp)) or T.I. ([takuhito.ito@riken.jp](mailto:takuhito.ito@riken.jp)).



## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Expression and purification of recombinant *S. pombe* eIF2B.** The detailed methods for expression and purification of the native eIF2B protein are described elsewhere (Kashiwagi *et al.*, submitted). For the production of the selenomethionine (SeMet)-substituted protein, the cells were grown in M9 medium until the absorbance,  $A_{600\text{ nm}}$ , reached 0.4, supplemented with SeMet and amino acids, grown for 30 min, induced with 0.5 mM IPTG, and further grown at 18 °C overnight<sup>30</sup>. The SeMet-derivative protein was also purified in the same manner as described, except the DTT concentration was increased to 10 mM during the purification and the Ni-Sepharose flow-through step was omitted. For cross-linking experiments, the epitope tags were fused to the regulatory subunits ( $\alpha$ : myc-tag at the N terminus;  $\beta$ : HA-tag at the N terminus;  $\delta$ : Strep-tag at the C terminus) using a PrimeSTAR Mutagenesis Basal Kit (Takara).

**Purification of *K. pastoris* eIF2.** We used *K. pastoris* eIF2 for nucleotide exchange assays and SEC analyses, because a high quality sample of *K. pastoris* eIF2 can be prepared more efficiently than *S. pombe* eIF2, by the following methods. First, the DNA sequence encoding the THF (Thrombin-His<sub>6</sub>-3 × Flag) tag was inserted into the *K. pastoris* genomic locus corresponding to the C terminus of eIF2 $\gamma$ , by homologous recombination<sup>31</sup>. The resultant *K. pastoris* strain producing the THF-tagged eIF2 $\gamma$  was grown in medium containing 1% yeast extract, 2% tryptone and 5% (v/v) glycerol at 30 °C. Harvested cells were suspended in 75 mM Tris-OAc buffer (pH 7.5), containing 300 mM KOAc, 10% (v/v) glycerol, 5 mM Mg(OAc)<sub>2</sub>, 1 mM EDTA, 2 mM DTT and protease inhibitors, and were disrupted using an FPG12800 Pressure Cell Homogenizer (Stansted Fluid Power). After centrifugation, the supernatant was purified by chromatography on Q-Sepharose, Ni-Sepharose, HiTrap Heparin, MonoS and Superdex 200 (GE Healthcare) columns. The final buffer was 20 mM HEPES-KOH buffer (pH 7.5), containing 150 mM KOAc, 5% (v/v) glycerol, 5 mM Mg(OAc)<sub>2</sub>, 0.1 mM EDTA and 1 mM DTT.

**Expression and purification of recombinant *S. pombe* eIF2 $\alpha$ .** *S. pombe* eIF2 $\alpha$  was produced as the N-terminally GST-tagged and C-terminally His<sub>6</sub>-Flag-tagged protein, in *E. coli* Rosetta2 (DE3). Transformed cells were grown in lysogeny broth (LB) medium at 37 °C. After the addition of 0.3 mM IPTG when the culture reached  $A_{600\text{ nm}} = 0.4$ , the cells were grown at 18 °C overnight. The cells were lysed by sonication in 20 mM HEPES-KOH buffer (pH 7.5), containing 150 mM KCl, 10% (v/v) glycerol, 1 mM DTT and protease inhibitors, and the lysate was fractionated on Ni-Sepharose and HiTrap Heparin columns. After overnight cleavage of the GST tag with the HRV 3C protease at 4 °C, eIF2 $\alpha$  was purified on the HiTrap Heparin and Superdex 200 columns in 20 mM HEPES-KOH buffer (pH 7.5), containing 150 mM KCl, 5% (v/v) glycerol and 1 mM DTT.

**In vitro phosphorylation of the  $\alpha$ -subunit of eIF2.** The active human PKR protein was prepared as previously described, with some modifications<sup>32,33</sup>. Briefly, the full-length PKR protein was expressed in *E. coli* Rosetta2 (DE3) cells as the N-terminally His<sub>6</sub>-tagged form, and dephosphorylated *in vivo* by co-expression with  $\lambda$  protein phosphatase. Transformed cells were grown in LB medium supplemented with 0.2% glucose at 37 °C. After the addition of 0.3 mM IPTG when the culture reached  $A_{600\text{ nm}} = 0.6$ , the cells were grown at 20 °C for 24 h. The cells were lysed by sonication in 20 mM HEPES-NaOH buffer (pH 7.5), containing 50 mM NaCl, 10% (v/v) glycerol, 1 mM DTT and protease inhibitors, and the lysate was centrifuged. The supernatant was fractionated on Heparin-Sepharose, Ni-Sepharose, HiTrap SP and Superdex 200 (GE Healthcare) columns. For the preparation of P-eIF2 $\alpha$  and eIF2( $\alpha$ P), PKR and *S. pombe* eIF2 $\alpha$  were dialysed against 20 mM HEPES-KOH buffer (pH 7.5), containing 50 mM KCl, 5 mM MgCl<sub>2</sub>, 0.1 mM EDTA and 1 mM DTT, and *K. pastoris* eIF2 was dialysed against the same buffer containing 100 mM KCl. PKR was concentrated to 2 mg ml<sup>-1</sup>, and activated by an incubation with 0.5 mM ATP at 30 °C for 1 h. eIF2 $\alpha$  and eIF2 were phosphorylated by an incubation with the activated PKR and 0.5 mM ATP at 25 °C. Phosphorylation was confirmed by Phos-tag SDS-PAGE (Wako).

**Guanine-nucleotide exchange assay.** The eIF2-[<sup>3</sup>H]GDP binary complex was formed by incubating 90 pmol of *K. pastoris* eIF2 with 37.5 pmol of [<sup>3</sup>H]GDP in assay buffer (20 mM HEPES-KOH buffer (pH 7.5) containing 100 mM KCl, 10% (v/v) glycerol, 0.1 mM EDTA, 5 mM NaF, 1 mM DTT and 2 mg ml<sup>-1</sup> BSA), at 21 °C for 10 min. After the addition of 3 mM MgCl<sub>2</sub> and 0.1 mM ATP, the eIF2-[<sup>3</sup>H]GDP complex was further incubated with or without 1  $\mu$ l of the activated PKR for 5 min, and kept on ice. After a 5 min incubation at 15 °C, measurements were started by the simultaneous additions of a 100-fold amount of GDP and 22.5 pmol of eIF2B. At each time point, samples were transferred into 2.5 ml of ice-cold wash buffer (20 mM HEPES-KOH buffer (pH 7.5) containing 100 mM KCl and 5 mM MgCl<sub>2</sub>), and immediately vacuum-filtered through mixed cellulose ester filters (Advantec). After two washes with 2.5 ml of the ice-cold wash buffer, the filters were dried and the radioactivity was quantitated by liquid scintillation counting.

**Size-exclusion chromatography analysis.** A mixture of 500 pmol of *S. pombe* eIF2B and 1 nmol of *K. pastoris* eIF2 was incubated in 20 mM HEPES-KOH buffer (pH 7.5), containing 300 mM KOAc, 5% (v/v) glycerol, 3 mM Mg(OAc)<sub>2</sub>, 0.1 mM EDTA and 1 mM DTT, at 4 °C. After the sample volume was adjusted to 200  $\mu$ l, it was fractionated on a Superose 6 column (GE Healthcare).

**Isothermal titration calorimetry.** *S. pombe* eIF2B and eIF2 $\alpha$  were applied to Sephacryl S-300 and Superdex 200 columns, respectively, in 20 mM HEPES-KOH buffer (pH 7.5), containing 300 mM KOAc, 5% (v/v) glycerol, 3 mM Mg(OAc)<sub>2</sub> and 0.1 mM EDTA. The measurements were performed with a MicroCal Auto-iTC200 calorimeter (Malvern). The titration was performed by injecting 2  $\mu$ l of eIF2 $\alpha$  (200  $\mu$ M) into the eIF2B solution (20  $\mu$ M), 19 times. To calculate the binding stoichiometry and the dissociation constants, two runs were concatenated.

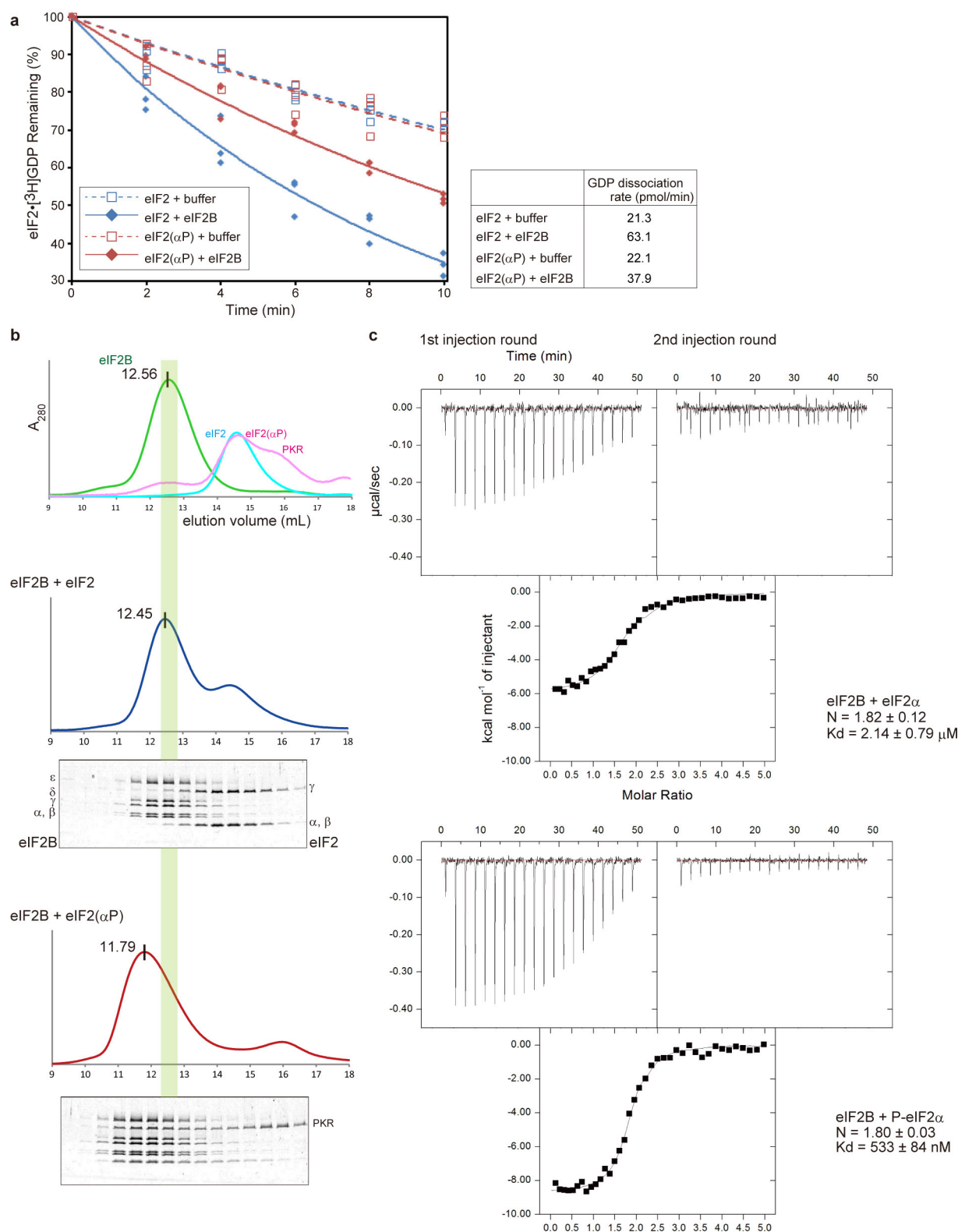
**Crystallization and structure determination.** The native and SeMet-derivative eIF2B samples were concentrated to 5–6 mg ml<sup>-1</sup>, and their crystals were obtained and cryoprotected in similar manners to those described in Kashiwagi *et al.* (submitted). The final data sets were collected at BL41XU of SPring-8 (Hyogo, Japan). Data collection was performed at 100 K, and the wavelengths were 1.0000 Å for the native data set and 0.9792 Å for the SeMet-derivative data set. The data sets were processed with XDS<sup>34</sup> and SCALA<sup>35</sup>. Data collection statistics are summarized in Extended Data Table 1. The initial phases were determined from the SeMet-derivative data set by the single-wavelength anomalous dispersion (SAD) method, by the use of the autoSHARP pipeline<sup>36</sup>. The molecular model was built automatically with Buccaneer<sup>37</sup>. The model was modified manually in Coot<sup>38</sup>, and refined by PHENIX<sup>39</sup> against the native data set. In the Ramachandran plot, 93.4% of the residues in the model are in the favoured region, and 6.0% and 0.6% of the residues are in the allowed and disallowed regions, respectively. The electrostatic surface potential was determined with APBS<sup>40</sup>.

**Surface-scanning site-directed photo-cross-linking assays.** The incorporation of pBpa was performed by the expanded genetic code method: pBpa was incorporated site-specifically into proteins using an *E. coli* RFzero strain, in which the UAG codon is reassigned to pBpa<sup>41,42</sup>. For the expression of pBpa-labelled *S. pombe* eIF2B, the codon at a specified position in the expression constructs (Kashiwagi *et al.*, submitted) was changed to a TAG triplet, using a PrimeSTAR Mutagenesis Basal Kit. For the pBpa-labelled *S. pombe* eIF2 $\alpha$ , the GST tag of the expression construct was replaced by MBP, and one codon was changed to TAG. The proteins labelled with pBpa were produced with the BL21 (DE3)-based RFzero strain, expressing pBpaRS and UAG-decoding tRNA<sup>41,42</sup>. The cells were grown in LB medium supplemented with 0.2% glucose and 1 mM pBpa, at 37 °C. When the culture reached  $A_{600\text{ nm}} = 0.8$ , 0.5 mM IPTG was added, and the cells were grown at 20 °C for 24 h. The pBpa-labelled eIF2B and eIF2 $\alpha$  were purified with Amylose Resin (New England Biolabs), and the MBP tags were cleaved with HRV 3C protease. For the cross-linking experiments between eIF2B and eIF2 $\alpha$ , the proteins were dialysed against 20 mM HEPES-KOH buffer (pH 7.5), containing 150 mM KCl, 5% (v/v) glycerol and 1 mM DTT. For the cross-linking experiments between eIF2B and *K. pastoris* eIF2, the proteins were dialysed against 20 mM HEPES-KOH buffer (pH 7.5), containing 150 mM KOAc, 5% (v/v) glycerol, 3 mM Mg(OAc)<sub>2</sub>, 0.1 mM EDTA and 1 mM DTT. The pBpa-labelled protein was divided into aliquots corresponding to the amount purified from 50 ml LB culture, and mixed with 100 pmol of the cross-linking target protein, and then the total volume was adjusted to 110  $\mu$ l. The final concentrations of the pBpa-labelled proteins were about 0.5–1.0  $\mu$ M. A series of protein variants, each with pBpa incorporated at one specified surface site, was ultraviolet-irradiated in the presence of the cross-linking target protein. Ultraviolet irradiation at 365 nm wavelength was performed on ice for 5 min. The cross-linked products were detected by western blotting. The Flag-tag, myc-tag, HA-tag and Strep-tag were detected using Anti-DDDDK-tag mAb-HRP-Direct (M185-7, MBL), Anti-c-Myc-Peroxidase antibody (A5598, Sigma), Monoclonal Anti-HA-Peroxidase antibody (H6533, Sigma) and Strep-Tag II Antibody HRP Conjugate (71591, Novagen), respectively. For the time course analysis of the cross-linking of the pBpa-labelled eIF2B with eIF2 or eIF2( $\alpha$ P), ATP alone or with the activated PKR was added to the dialysed eIF2, kept at 21 °C for 5 min and then mixed with eIF2B.

- Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
- Higo, T. *et al.* Development of a hexahistidine-3 × FLAG-tandem affinity purification method for endogenous protein complexes in *Pichia pastoris*. *J. Struct. Funct. Genomics* **15**, 191–199 (2014).
- Matsui, T., Tanihara, K. & Date, T. Expression of unphosphorylated form of human double-stranded RNA-activated protein kinase in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **284**, 798–807 (2001).
- Lemaire, P. A., Lary, J. & Cole, J. L. Mechanism of PKR activation: dimerization and kinase activation in the absence of double-stranded RNA. *J. Mol. Biol.* **345**, 81–90 (2005).
- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).

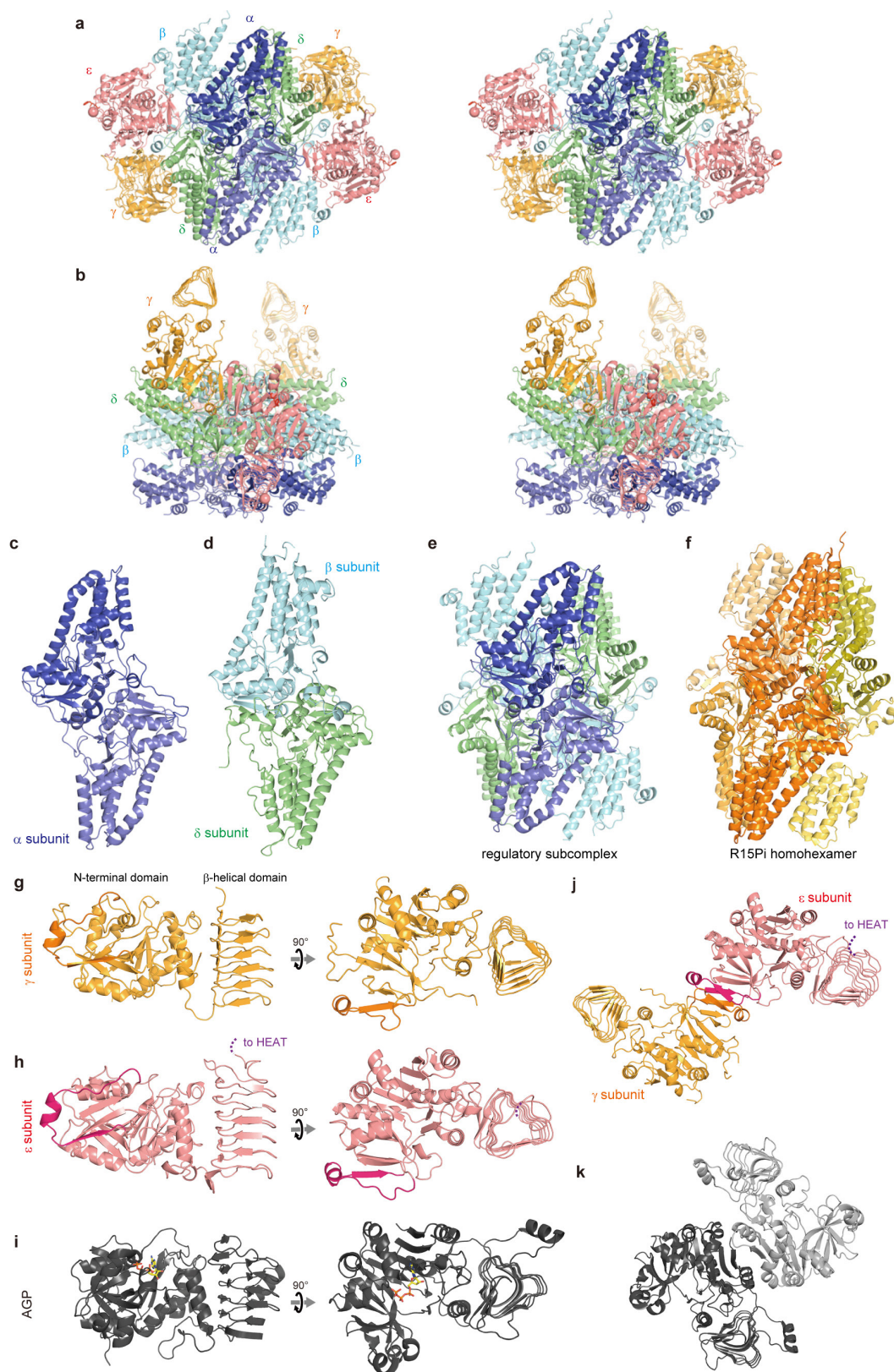


35. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
36. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
37. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
39. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
40. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
41. Mukai, T. *et al.* Genetic-code evolution for protein synthesis with non-natural amino acids. *Biochem. Biophys. Res. Commun.* **411**, 757–761 (2011).
42. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 11020–11024 (2002).
43. Hiyama, T. B., Ito, T., Imataka, H. & Yokoyama, S. Crystal structure of the  $\alpha$  subunit of human translation initiation factor 2B. *J. Mol. Biol.* **392**, 937–951 (2009).
44. Nakamura, A. *et al.* Dynamic, ligand-dependent conformational change triggers reaction of ribose-1,5-bisphosphate isomerase from *Thermococcus kodakarensis* KOD1. *J. Biol. Chem.* **287**, 20784–20796 (2012).
45. Jin, X., Ballicora, M. A., Preiss, J. & Geiger, J. H. Crystal structure of potato tuber ADP-glucose pyrophosphorylase. *EMBO J.* **24**, 694–704 (2005).
46. Geva, M. *et al.* A mouse model for eukaryotic translation initiation factor 2B-leucodystrophy reveals abnormal development of brain white matter. *Brain* **133**, 2448–2461 (2010).
47. Dev, K. *et al.* Archaeal eIF2B interacts with eukaryotic translation initiation factors eIF2 $\alpha$  and eIF2B $\alpha$ : implications for eIF2B function and eIF2B regulation. *J. Mol. Biol.* **392**, 701–722 (2009).



**Extended Data Figure 1 | Biochemical properties of recombinant *S. pombe* eIF2B.** **a**, Guanine-nucleotide exchange catalysed by *S. pombe* eIF2B, and its inhibition by eIF2 phosphorylation. *K. pastoris* eIF2 was labelled with [ $^3\text{H}$ ]GDP and incubated with ATP and activated PKR (eIF2( $\alpha$ P); red line), or with ATP only (eIF2; blue line). Reactions were started by the addition of excess unlabelled GDP with eIF2B (solid line) or buffer (dashed line). Individual data points of triplicate analyses are shown by dots and boxes. The dissociation rates of GDP from eIF2 are summarized in the table. **b**, The SEC profiles of respective proteins (*S. pombe* eIF2B (green), *K. pastoris* eIF2 (cyan), eIF2( $\alpha$ P) (pink)) (top),

eIF2B with *K. pastoris* eIF2 (middle) and eIF2B with eIF2( $\alpha$ P) (bottom). The chromatograms of the absorbance at 280 nm and the Coomassie blue stained SDS-PAGE gels are shown. For gel source data, see Supplementary Fig. 1. **c**, ITC measurements between *S. pombe* eIF2B (as the decamer) and *S. pombe* eIF2 $\alpha$  (top), eIF2B and P-eIF2 $\alpha$  (bottom). We used the eIF2 $\alpha$  subunit alone, rather than the trimeric eIF2, in these experiments because of the difficulty in the preparation of highly concentrated eIF2. Representative thermograms are shown. Two runs were concatenated for analysis. The binding stoichiometry ( $N$ ) and the dissociation constant ( $K_d$ ) were calculated from triplicate analyses (mean  $\pm$  s.d.).



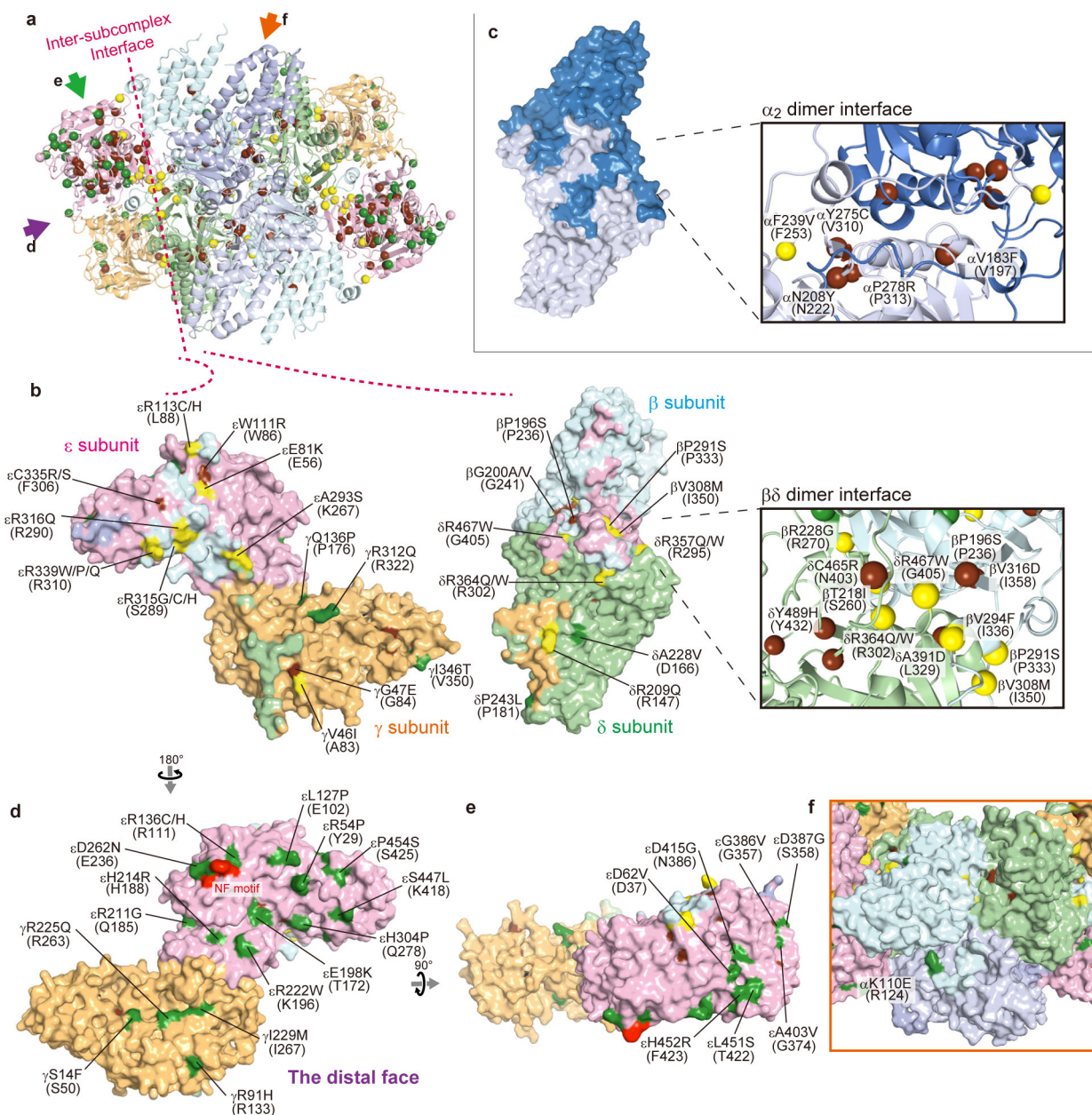
Extended Data Figure 2 | See next page for figure caption.



**Extended Data Figure 2 | Architecture of the eIF2B subcomplexes.**

**a, b**, Two different views of the crystal structure of *S. pombe* eIF2B (wall-eyed stereo view), coloured as in Fig. 1a and from the same view as in Fig. 1b and Fig. 1c, respectively. The assembly of the subcomplexes is primarily mediated by the  $\beta$ - $\epsilon$  and  $\delta$ - $\gamma$  interactions. In this regard, the results of the two prior mutational analyses<sup>12,13</sup> are consistent with the present structure, except that one suggested that the  $\beta$ -helical region of the  $\gamma$ -subunit is involved in the inter-subunit interactions. **c-k**, Ribbon models of the structures of *S. pombe* eIF2B subcomplexes (**c-e**, **g**, **h**, **j**) and proteins with structural similarity (**f**, **i**, **k**). **c, d**, The  $\alpha_2$  homodimer (**c**) and the  $\beta\delta$  heterodimer (**d**) in the eIF2B decamer. The conformations of each subunit are similar to those in the human  $\alpha_2$  homodimer<sup>43</sup> and the *Chaetomium thermophilum*  $\beta\delta$  heterotetramer<sup>3</sup>, even though the relative orientations of the subunits in the dimers are slightly different from those represented in these partial structures. **e, f**, The regulatory subcomplex in the eIF2B

decamer (**e**) and ribose-1,5-bisphosphate (R15Pi) homohexamer<sup>44</sup> (PDB 3A11) (**f**). The architecture of the regulatory subcomplex is an assembly of three similarly shaped dimer moieties: one homodimer of the  $\alpha$ -subunit and two heterodimers of the  $\beta$ - and  $\delta$ -subunits. The arrangement of the regulatory subunits resembles that in the *C. thermophilum*  $\beta\delta$  heterotetramer<sup>3</sup>, and shares some similarity to that in the homohexameric structure of R15Pi<sup>44</sup>. **g**, The  $\gamma$ -subunit of eIF2B. **h**, The  $\epsilon$ -subunit of eIF2B. **i**, Potato tuber ADP-glucose pyrophosphorylase (AGP) (PDB 1YP3)<sup>45</sup>. The dimerization interfaces between the catalytic subunits are coloured in deeper shades (**g**, **h**). **j, k**, The subunit heterodimerization mode in the catalytic subcomplex of eIF2B (**j**) and the subunit homodimerization mode in the potato tuber AGP (**k**). The dimerization manner of the  $\gamma$ - and  $\epsilon$ -subunits is novel: each of their structures resembles the subunit structure of the AGP homotetramer<sup>45</sup>, but they dimerize through their N-terminal domains, in a different manner than the AGP homotetramer<sup>45</sup>.

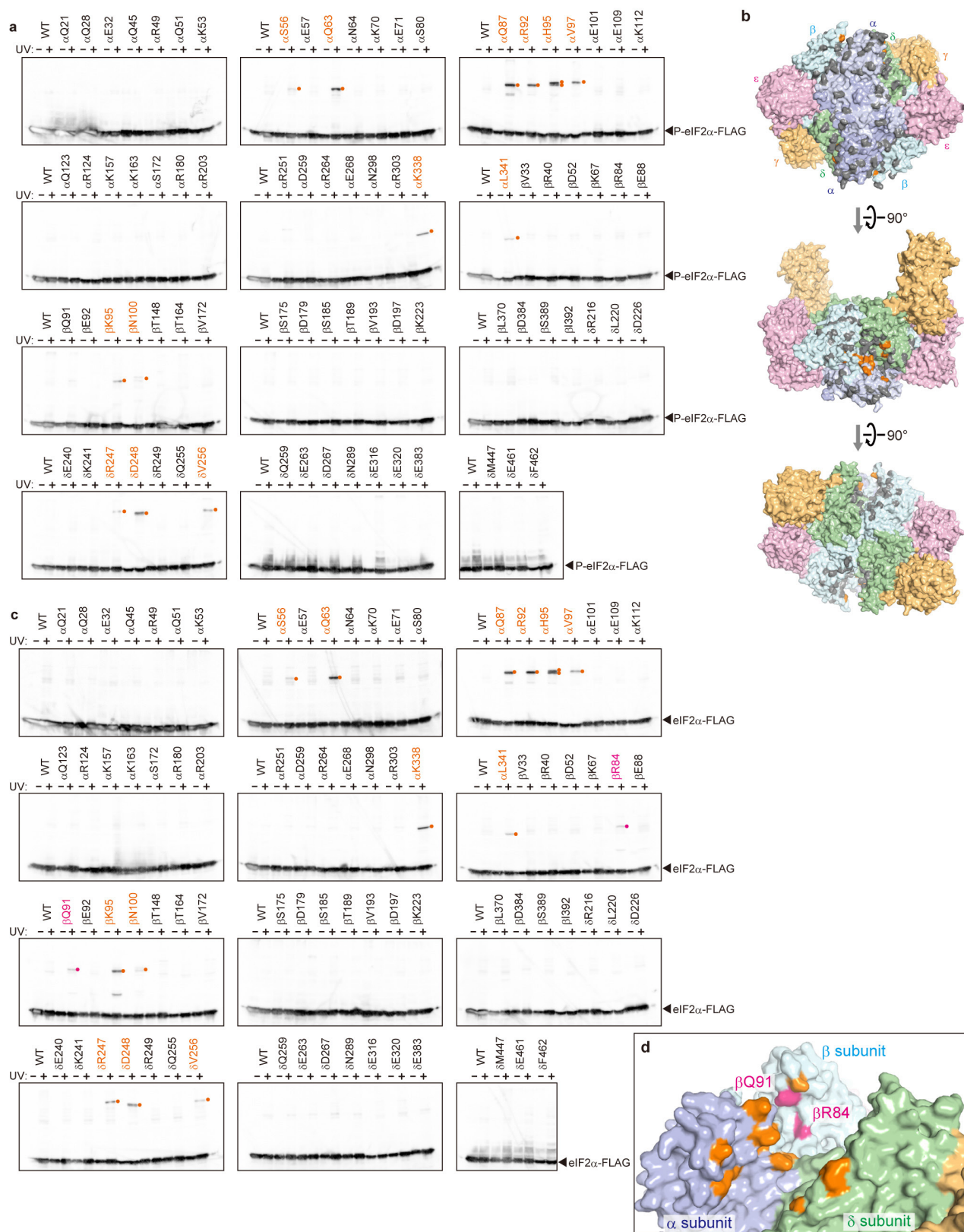


**Extended Data Figure 3 | Mapping of the residues corresponding to missense VWM mutations on the subunit interfaces and the distal face of eIF2B.** The eIF2B residues corresponding to VWM-causing missense mutations in human (Supplementary Table 1) are mapped on the *S. pombe* eIF2B structure, with the same subunit colouring as in Fig. 1. The *S. pombe* eIF2B residues corresponding to VWM-causing missense mutations are shown in parentheses. **a**, VWM-related residues are mapped as spheres on the overall structure (ribbon model). The environments of the residues are colour-coded (green, solvent-exposed; yellow, subunit interface; brown, structural core) on the spheres. **b**, VWM-related residues are mapped on the surface model of the inter-subcomplex interfaces on the catalytic subcomplex side (left), with the interfaces for the  $\alpha$ -,  $\beta$ - and  $\delta$ -subunits coloured blue, cyan and green, respectively, and on the  $\beta\delta$  dimer side (right), with the interfaces for the  $\gamma$ - and  $\epsilon$ -subunits coloured orange and pink, respectively. VWM-related residues around the  $\beta\delta$  dimerization interface are shown as spheres in the inset. The mutations in the regulatory subunits are clustered around the dimerization interface between the  $\beta$ - and  $\delta$ -subunits, as mentioned in ref. 3. Our structure

further revealed that the binding site for the  $\epsilon$ -subunit is formed by the correct interaction between the  $\beta$ - and  $\delta$ -subunits, thus explaining the abundance of mutations around this interface. **c**, VWM-related residues in the  $\alpha_2$  homodimer are located around the homodimerization interface, and shown as spheres in the inset. These VWM-related mutations around the subunit interfaces (**b**, **c**) may cause appreciable degrees of subunit dissociation from the eIF2B decamer, leading to incomplete complexes, destabilization of eIF2B resulting in aggregation/degradation, and/or changes in the conformation and activity of the intact eIF2B decamer. **d**, VWM-related residues on the distal face of the catalytic subcomplex are mapped on the surface model. The NF motif is shown in red. Several VWM-related residues are located near the NF motif, including Arg111(2B $\epsilon$ ), corresponding to the human Arg136His(2B $\epsilon$ ) mutation, for which the mouse model is available<sup>46</sup>. **e**, The  $\epsilon$ -subunit further contains several exposed missense VWM mutations, especially in the  $\beta$ -helical domain. **f**, VWM-related residues in the central cavity. Only one residue, corresponding to the human Lys110Glu(2B $\alpha$ ) mutation, is exposed to the solvent.

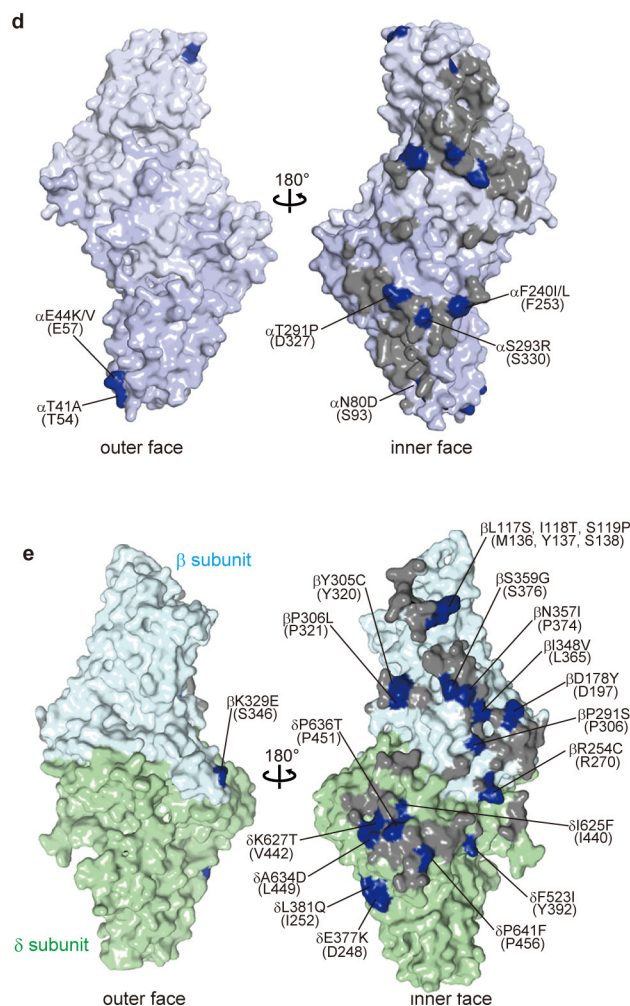
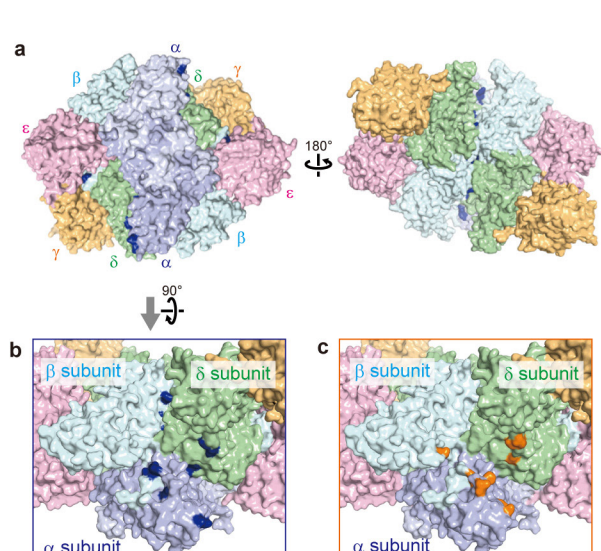






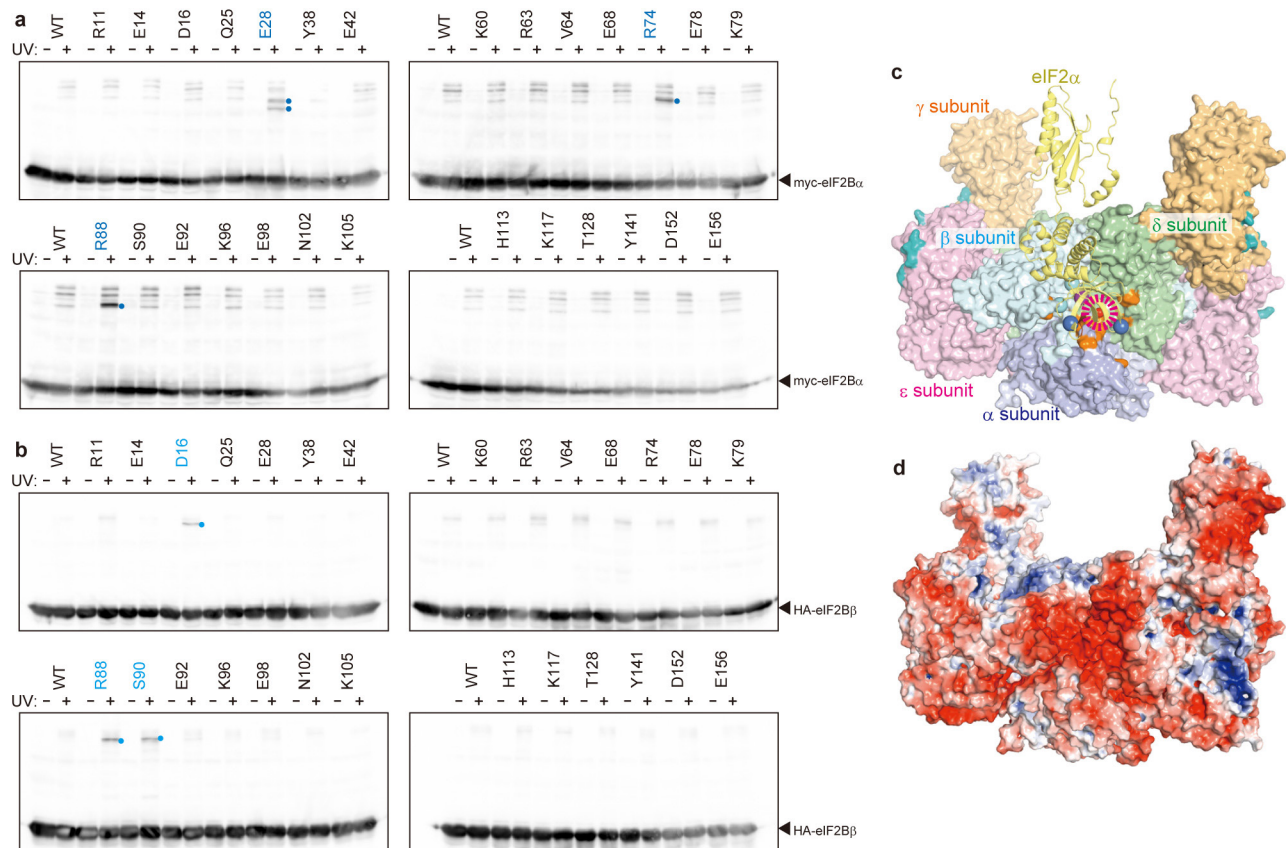
**Extended Data Figure 5 | Photo-cross-linking between pBpa-labelled eIF2B and eIF2 $\alpha$ .** **a**, The *S. pombe* eIF2B variants bearing a single site-specific pBpa substitution in the regulatory subunit were mixed with *S. pombe* P-eIF2 $\alpha$ , and irradiated with ultraviolet (365 nm) for 5 min. Cross-linking was detected as described in Extended Data Fig. 4. The relevant bands are indicated with orange dots. **b**, The eIF2 $\alpha$ -cross-linked sites are shown in orange, and the cross-link-negative sites are shown in grey, on the surface model of the overall structure of eIF2B, with the

same subunit colouring as in Fig. 1. **c**, The eIF2B variants were similarly mixed with unphosphorylated eIF2 $\alpha$  and irradiated with ultraviolet. The bands that were also observed in **a** are indicated with orange dots, and the unphosphorylated eIF2 $\alpha$ -specific bands are indicated with magenta dots. **d**, Arg84(2B $\beta$ ) and Gln91(2B $\beta$ ), which exhibited unphosphorylated eIF2 $\alpha$ -specific cross-links, are shown in magenta. The view is the same as in Fig. 2a. For gel source data, see Supplementary Fig. 1.



**Extended Data Figure 6 | Solvent-exposed Gcn<sup>-</sup> mutations located in the central cavity and buried Gcn<sup>-</sup> mutations clustered around the trimerization interface.** **a**, The residues corresponding to Gcn<sup>-</sup> mutations<sup>19,20</sup> are mapped in blue on the surface model of the *S. pombe* eIF2B structure, with the same subunit colouring as in Fig. 1. **b**, **c**, The residues corresponding to exposed Gcn<sup>-</sup> mutations are mainly located in the central cavity (**b**), and their locations coincide with the P-eIF2 $\alpha$  cross-link sites shown in orange (**c**, Extended Data Fig. 5b). **d**, **e**, The residues corresponding to Gcn<sup>-</sup> mutations on the  $\alpha_2$  homodimer (**d**) and

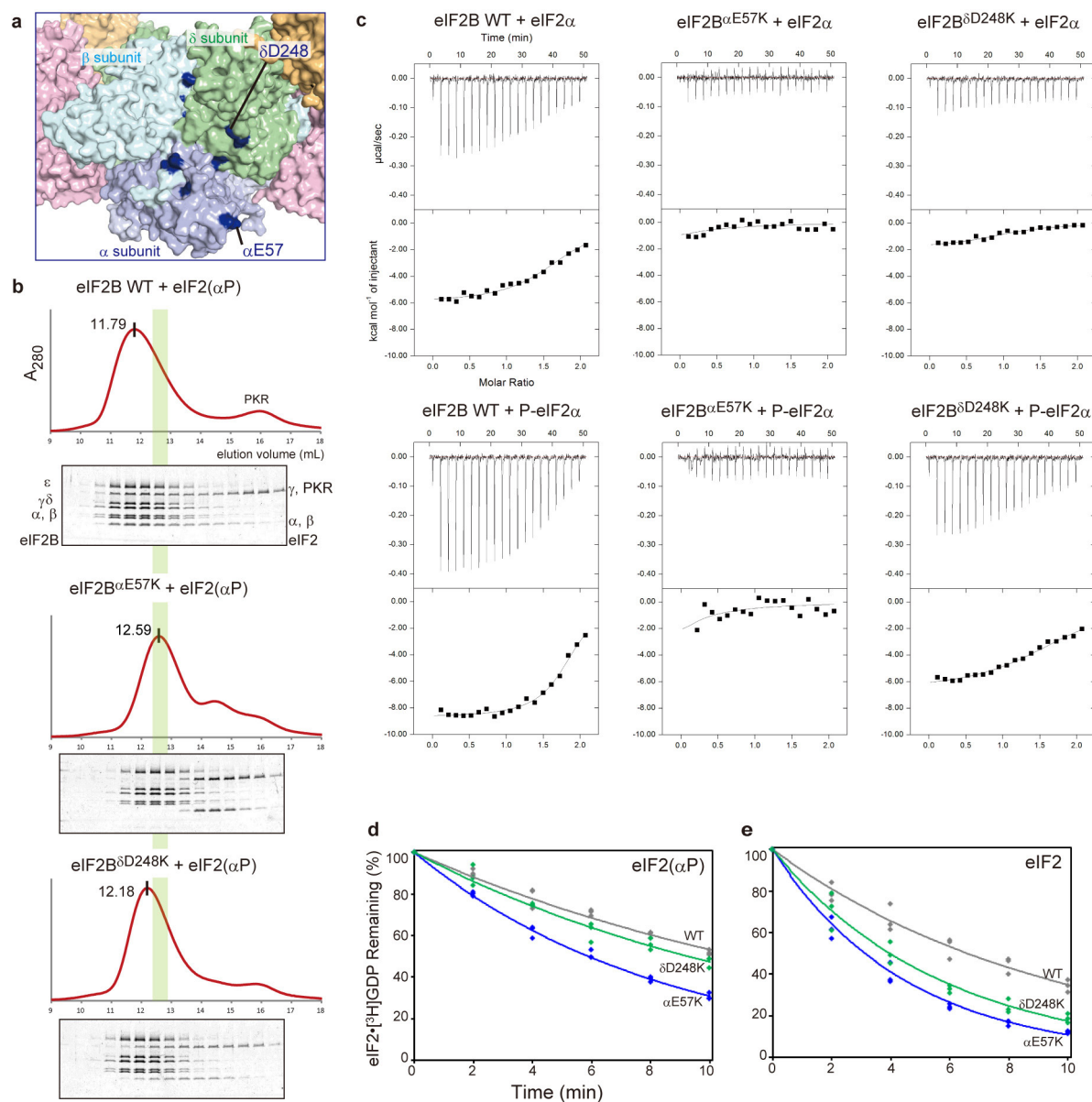
the  $\beta\delta$  heterodimer (**e**). The residues of *S. pombe* eIF2B corresponding to *S. cerevisiae* Gcn<sup>-</sup> mutations are indicated in parentheses. The interfaces for the trimerization of the regulatory dimers are coloured grey. Most of the Gcn<sup>-</sup>-related residues are mapped only on one face of the subunits, as predicted<sup>43,47</sup>. The present eIF2B structure revealed that these 'mutation-rich' faces are used for the assembly of the regulatory subunits to form the subcomplex, and the Gcn<sup>-</sup>-related residues are clustered on the interface for the trimeric assembly.



**Extended Data Figure 7 | Cross-linking of pBpa-labelled eIF2 $\alpha$  with the central cavity formed by the eIF2B regulatory subunits. a, b,** The *S. pombe* eIF2 $\alpha$  variants bearing a single site-specific pBpa residue were mixed with epitope-tagged *S. pombe* eIF2B, and irradiated with ultraviolet (365 nm) for 5 min on ice. Cross-linking was detected as described in Extended Data Fig. 4. Cross-links with eIF2B $\alpha$  were detected with an anti-myc antibody (**a**), and those with eIF2B $\beta$  were detected with an anti-HA antibody (**b**). The relevant bands are indicated with blue dots in **a**, and cyan dots in **b**. These cross-linked sites are mapped on the human eIF2 $\alpha$  structure<sup>21</sup> in Fig. 2b. For gel source data, see Supplementary Fig. 1. **c**, The model of the eIF2B–eIF2 $\alpha$  complex, built on the basis of the cross-linking experiments, is shown in Fig. 2c. The phosphorylated residue Ser51(2 $\alpha$ ) is highlighted with the magenta circle. **d**, The electrostatic surface

potential of eIF2B, from the same viewpoint as in **c**. Red and blue colours represent negative and positive potentials, respectively, of  $\pm 10kT/e$ . The cavity has no positively charged patch; therefore, the mechanism underlying the enhanced affinity for P-eIF2 $\alpha$  is still unclear. One possible mechanism is a cation-mediated recognition of the phosphoserine residue. The phosphorylated Ser51(2 $\alpha$ ) may coordinate a cation together with the negatively charged residues at the bottom of the central cavity, although we did not observe any electron density for such cations in the cavity. Another possibility is a phosphorylation-induced conformational change of the Ser51-flanking loop. The phosphorylation of Ser51(2 $\alpha$ ) may induce the rearrangement of adjacent arginine residues, and enable a stronger interaction with the negatively charged residues at the bottom of the central cavity.

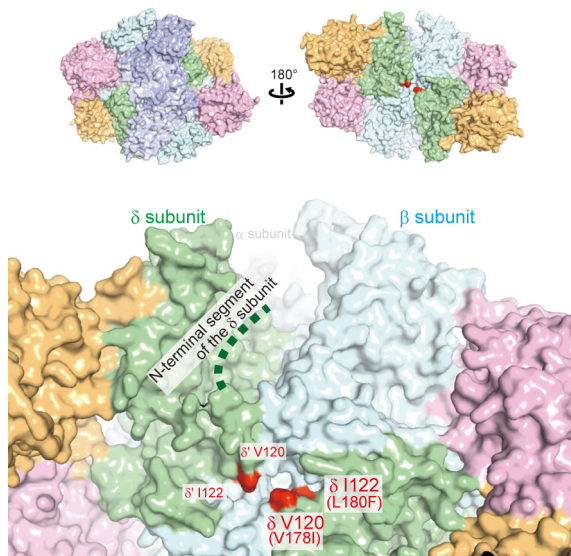




#### Extended Data Figure 8 | Analyses of eIF2B Gcn<sup>-</sup> mutations.

**a**, The locations of the Gcn<sup>-</sup>-related residues mutated for the SEC and ITC analyses (Glu57(2B $\alpha$ ) and Asp248(2B $\delta$ )) are indicated on the *S. pombe* eIF2B structure, with the same view and colouring as in Extended Data Fig. 6b. **b**, The SEC analyses of the interaction between *K. pastoris* eIF2( $\alpha$ P) and *S. pombe* eIF2B, bearing Gcn<sup>-</sup>-related mutations. The chromatograms of the absorbance at 280 nm and the SDS-PAGE gels of each run are shown. The green bar represents the elution point of free

eIF2B (Extended Data Fig. 1b). For gel source data, see Supplementary Fig. 1. **c**, ITC measurements between *S. pombe* eIF2 $\alpha$  or P-eIF2 $\alpha$  and eIF2B, bearing Gcn<sup>-</sup>-related mutations. Representative thermograms for the ITC experiments are shown. **d**, **e**, The nucleotide exchange activities of *S. pombe* eIF2B, bearing Gcn<sup>-</sup>-related mutations, on *K. pastoris* eIF2( $\alpha$ P) (**d**) and eIF2 (**e**) were examined as described in Extended Data Fig. 1a ( $\alpha$ Glu57Lys mutant, blue line;  $\delta$ Asp248Lys mutant, green line; wild type, grey line). Individual data points of triplicate analyses are shown by dots.



**Extended Data Figure 9 | The locations of ISRIB-resistant mutations on the eIF2B structure.** Residues corresponding to the ISRIB-resistant mutations<sup>26</sup> are mapped onto the eIF2B structure in red. The residue corresponding to the Arg171Gln(2Bδ) mutation (Lys112(2Bδ)) is in the disordered region, at the N terminus of the δ-subunit. The disordered N-terminal segment of the δ-subunit is indicated by the dotted green line.

Extended Data Table 1 | Data collection and refinement statistics

	Native	SeMet
<b>Data collection</b>		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	144.5, 209.2, 223.5	144.6, 209.8, 223.8
Resolution (Å)*	49.29–2.99 (3.17–2.99)	44.25–3.38 (3.57–3.38)
$R_{\text{sym}}$ *	0.140 (1.558)	0.137 (0.803)
$I/\sigma I$ *	13.3 (1.3)	20.7 (3.7)
Completeness (%)*	99.6 (97.9)	99.8 (98.9)
Redundancy*	6.3 (6.3)	14.9 (14.7)
$CC_{1/2}$ *	0.998 (0.524)	
<b>Refinement</b>		
Resolution (Å)	44.29–2.99	
No. reflections	136110	
$R_{\text{work}}/R_{\text{free}}$	0.221 / 0.271	
No. atoms		
Protein	28539	
Ligand/ion	45	
Water	0	
B-factors		
Protein	86.45	
Ligand/ion	86.54	
R.m.s. deviations		
Bond lengths (Å)	0.007	
Bond angles (°)	1.44	

\*Highest resolution shell is shown in parentheses.



# CORRECTIONS & AMENDMENTS

---

## CORRIGENDUM

doi:10.1038/nature16180

### **Corrigendum: Acute stress facilitates long-lasting changes in cholinergic gene expression**

Daniela Kaufer, Alon Friedman, Shlomo Seidman & Hermona Soreq

*Nature* **393**, 373–377 (1998); doi:10.1038/30741

It has been brought to our attention that Figs 1a and 5a of this Letter contain some irregularities in the data presentation. Specifically, in the three gel images in the top row and in the rightmost gel image in the second row of Fig. 1a, a black area of the gel was cropped and added to unify the size of the box in order to match the size of the gel images without rescaling. All bands originated from the same gel. The histograms in the right column show quantification of the data from multiple experiments. In Fig. 5a, the similarity between the left and right gel images in the bottom row may be due to accidental duplication. The fact that a long time has passed since publication means it is no longer possible to obtain and investigate the raw data. None of these irregularities affect the original meaning of the experiments, their results, their interpretation, or the conclusions of the paper. Author S.S. is deceased.

Correspondence should be addressed to D.K. (danielak@berkeley.edu), A.F. (alonf@bgu.ac.il), and H.S. (hermona.soreq@mail.huji.ac.il).

# CORRECTIONS & AMENDMENTS

## CORRIGENDUM

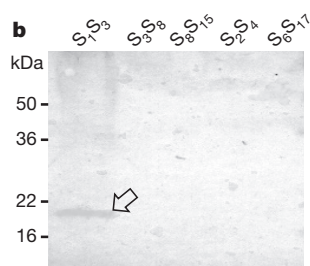
doi:10.1038/nature16181

### Corrigendum: Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*

Michael J. Wheeler, Barend H. J. de Graaf, Natalie Hadjosif, Ruth M. Perry, Natalie S. Poulter, Kim Osman, Sabina Vatovec, Andrea Harper, F. Christopher, H. Franklin & Veronica E. Franklin-Tong

*Nature* **459**, 992–995 (2009); doi:10.1038/nature08027

Recently, it has come to our attention that in the left panel of Fig. 2b of this Letter, the lanes labelled  $S_2S_4$  and  $S_6S_{17}$  were duplicated. We have reviewed the original data. It seems likely that a duplicated part of the blot was placed over lane  $S_6S_{17}$  to aid alignment of molecular mass markers and inadvertently left there. We have now removed the duplicated lane and show the whole western blot (Fig. 1). Our conclusions are unaffected.



**Figure 1** | This is the corrected left panel of Fig. 2b. The right panel showing the Coomassie staining is not shown.

## CORRIGENDUM

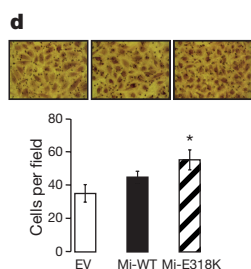
doi:10.1038/nature16158

### Corrigendum: A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma

Corine Bertolotto, Fabienne Lesueur, Sandy Giuliano, Thomas Strub, Mahaut de Lichy, Karine Bille, Philippe Dessen, Benoit d'Hayer, Hamida Mohamdi, Audrey Remenieras, Eve Maubec, Arnaud de la Fouchardière, Vincent Molinié, Pierre Vabres, Stéphane Dalle, Nicolas Poulalhon, Tanguy Martin-Denavit, Luc Thomas, Pascale Andry-Benzaquen, Nicolas Dupin, Françoise Boitier, Annick Rossi, Jean-Luc Perrot, Bruno Labeille, Caroline Robert, Bernard Escudier, Olivier Caron, Laurence Brugières, Simon Saule, Betty Gardie, Sophie Gad, Stéphane Richard, Jérôme Couturier, Bin Tean Teh, Paola Ghiorzo, Lorenza Pastorino, Susana Puig, Celia Badenas, Hakan Olsson, Christian Ingvar, Etienne Rouleau, Rosette Lidereau, Philippe Bahadoran, Philippe Vielh, Eve Corda, Hélène Blanché, Diana Zelenika, Pilar Galan, The French Familial Melanoma Study Group, Valérie Chaudru, Gilbert M. Lenoir, Mark Lathrop, Irwin Davidson, Marie-Françoise Avril, Florence Demenais, Robert Ballotti & Brigitte Bressac-de Paillerets

*Nature* **480**, 94–98 (2011); doi:10.1038/nature10539

In this Letter, one image was mistakenly duplicated during preparation of the artwork. In the original Fig. 3d, the left image illustrating migration of RCC4 cells transduced with empty adenovirus (EV) at 24 h is a duplicate of the middle image showing migration of RCC4 cells transduced with an adenovirus encoding Mi-WT. The corrected images and migration graph are shown in Fig. 1 of this Corrigendum. This correction does not alter any of the conclusions, and the authors apologize for any confusion this may have caused. *Nature* has not received a response from the following authors to approve this Corrigendum: V.M., T.M.-D., A.Ro., P.B., E.C. and V.C., and C. Becuwe., J.-L.B., J.C.-B., S.D., C.D., J.L., and K.M. from The French Familial Melanoma Study Group (L.D. is deceased).



**Figure 1** | This is the corrected Fig. 3d of the original Letter.

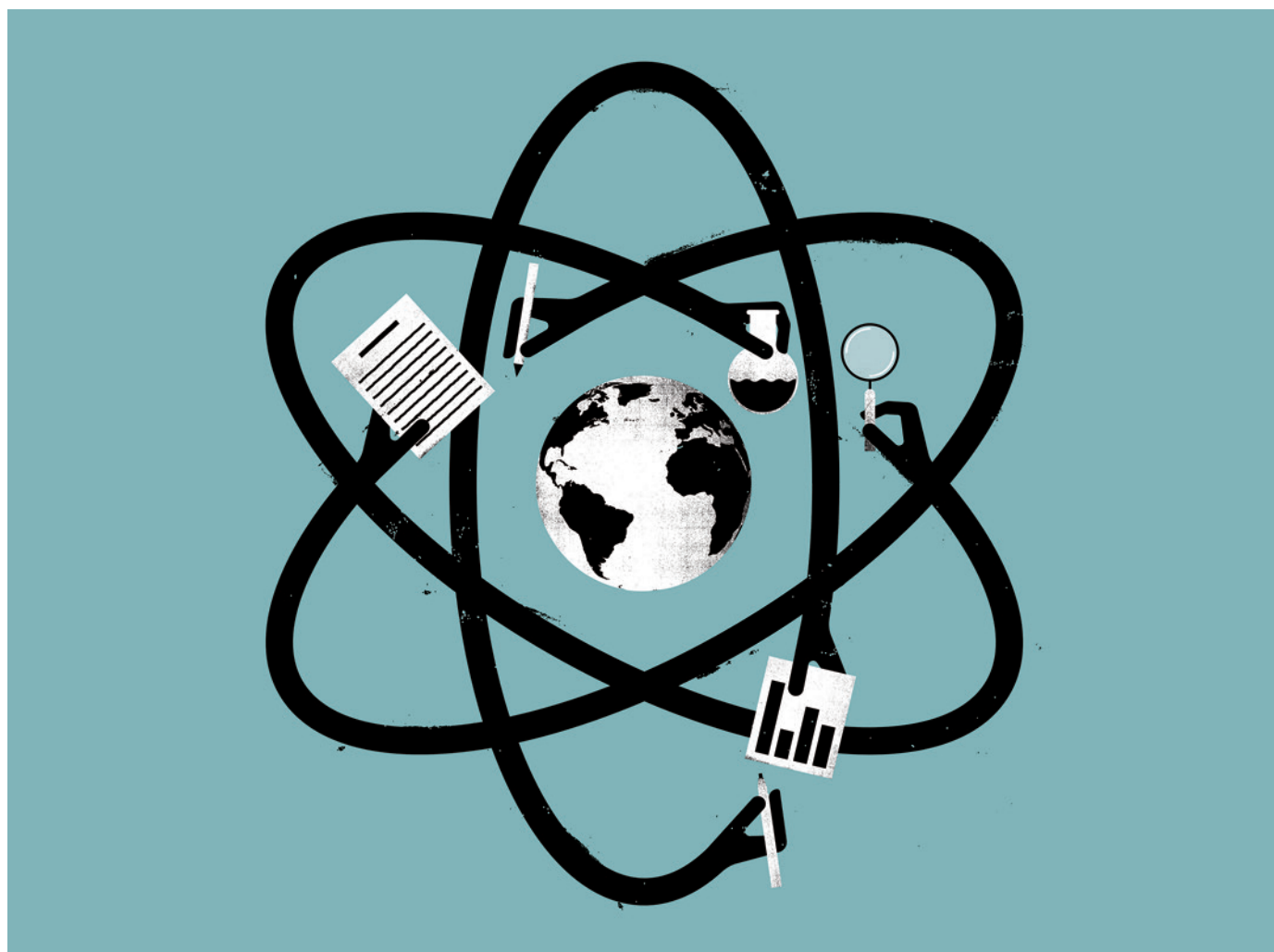


TOOLBOX

# THE MANUSCRIPT-EDITING MARKETPLACE

*A peer-to-peer website aims to disrupt the author-services industry.*

ILLUSTRATION BY THE PROJECT TWINS



BY JEFFREY M. PERKEL

As Sebastian Eggert prepared to submit a conference article, he realized he had a problem: neither he nor his research adviser were native English speakers, and neither had much experience in writing and publishing research papers. But Eggert, a master's student in mechanical engineering at the Technical University of Munich in Germany, had heard of a website where he could purchase editing services from an

expert: an online marketplace called Peerwith.

Launched in October 2015 and still in beta testing, Peerwith is a forum through which researchers can find and negotiate with service providers such as editors, translators, statisticians and illustrators to improve their research papers. The site boasts "hundreds of experts", most of them with expertise in the social sciences and humanities. Users post a job request detailing the subject area of the document, its length and the desired turnaround time. Experts then bid for the job, and both experts and users

rate each other afterwards. Peerwith's business model is akin to freelance marketplaces such as Upwork, says co-founder Joris van Rossum, who left the journal publisher Elsevier to start his firm, except with a strictly academic focus.

A market for author services on research papers already exists; van Rossum estimates it at hundreds of million of dollars annually. It includes both large editing companies such as American Journal Editors (AJE), Edanz, Editage and Macmillan Science Communication (MSC, which is owned by *Nature's* parent company),

and freelancers. But a peer-to-peer online marketplace, van Rossum says, makes services more affordable by cutting out the middleman and efficiently matching buyers and sellers. (Peerwith receives a cut of 10–20% for each transaction; the other firms would not comment on their margins). At the site, authors can review the experts who bid for work to identify the best fit, and can check to see how others have rated them.

Val Kidd, an editor and translator based in the United Kingdom, earned €200 (US\$223) on Peerwith to translate a presentation for Emanuel Rutten, a philosopher at the Free University, Amsterdam, in the Netherlands. The process, from job posting to completed document, took less than two weeks, Rutten says. “It’s really smooth.” For her part, Kidd says that the interaction with her client improved the final product. At most author-services companies Kidd works with, she says, editors and translators cannot contact the author should they have questions — the client interacts with the service, which identifies a freelancer to handle the job.

Peerwith doesn’t vet its service providers, says Anna Sharman, founder of Cofactor, a London-based author-services consultancy. So, unlike her own and other such companies, there is no guarantee that the ‘experts’ really are qualified. Editors at Cofactor undergo a rigorous recruitment process, Sharman says, and she double-checks their work before it is returned to the client.

Sharman says that she could see Peerwith as a marketing channel for her business, but is concerned that it may foster a “race to the bottom” in pricing. She says that when she created an account, the only request she saw was from someone who wanted a 5,000-word article edited for US\$9, “a ridiculously small amount”. Sharman charges £60 (\$87) per 1,000 words at Cofactor. At Editage, a 6,000-word article with 1-week turnaround costs \$350 at the company’s ‘premium’ price, and AJE charges \$594. And for ‘extensive’ scientific editing at MSC (by a panel of at least four editors with experience at high-impact journals), a typical 5,000-word article with a 17-day turnaround costs \$2,860.

Peerwith is still getting up to speed, van Rossum says. But ultimately, a community-based marketplace could succeed “if there’s the right balance of price and quality”, says Deni Auclair, a lead analyst for the media, information and technology consulting firm Outsell, headquartered in Burlingame, California. The larger editorial service providers might be left to target institutions more than individuals, she suggests.

As for Eggert, he received one bid to his job posting, and paid €100 for style and content edits to his 2,500-word paper, which he negotiated down from €120. He says he would use the service again, and recommend it to others — assuming the price is right. ■

## EXPERIMENTATION

# Better designs for animal studies

*A web application aims to improve life–sciences research.*

BY DANIEL CRESSEY

A free online tool that visualizes the design of animal experiments and gives critical feedback could save scientists from embarking on poorly designed research, the software’s developers hope.

Over the past few years, researchers have picked out numerous flaws in the design and reporting of published animal experiments, which, they warn, could lead to bias. In response, hundreds of journals have agreed to voluntary guidelines for reporting animal studies: checklists of best practice, such as what statistical calculations to use to ward off error.

But these lists kick in after scientists submit a paper, says Nathalie Percie du Sert, who specializes in experimental design at the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) in London. “When you get to the reporting stage, that’s a bit too late,” she says. “We want researchers to think about these issues at the design stage.”

Percie du Sert’s solution is a programme called the Experimental Design Assistant (EDA), which launched in October 2015. She hopes that it will help to improve the quality of animal research and perhaps even become an integral part of the conduct of animal studies.

The EDA ([go.nature.com/koasai](http://go.nature.com/koasai)) allows scientists to create a visual representation of an experiment by laying out its key elements — hypothesis, experimental method and planned analysis — in logically connected, coloured boxes. The software then uses a built-in set of rules to spot potential problems, and suggests refinements. These may be simple — the researcher hasn’t specified how to randomize animals to the control or treatment arm — or more complex: there are potential confounding variables in the control and trial arms. The tool can also assist scientists with calculating the sample size needed to ensure a statistically robust result, or with randomization.

There’s nothing fundamentally new in the EDA, says Percie du Sert. It builds on existing knowledge of good experimental design. But it can aid scientists who have little training in the area, she says, and teach them design choices.

Since the EDA’s launch, around 400 accounts have been created to use it, producing between 50 and 100 experimental diagrams in total each month, says Percie du Sert. She does not have access to detailed information about its users; the sensitivities around animal research and the need to protect researchers mean that data on who is using it, and how, are secured.

The Wellcome Trust’s Sanger Institute, a genome-research centre in Cambridge, UK, is rolling out an internal training programme that includes lessons on design and use of the EDA; the agency is encouraging staff to use the software to present experiments to ethical-review committees, says Natasha Karp, a biostatistician at the institute. Karp took part in the working group that oversaw the tool’s development, and says that she uses it to visualize the experiments of the biologists whom she supports.

The EDA is not the only software that aims to improve research quality and reproducibility. Other tools check manuscripts before publication for issues such as errors in formatting or omission of *P* values. These include Penelope, a paid-for service aimed at journal publishers; another tool called WebCONSORT (which is not yet freely available) is being tested as a way to improve reporting of clinical trials. Protocol Navigator, a free web application created by scientists at Cardiff University, UK, also produces visual experiment maps that can be shared.

But the EDA specifically targets animal research, and as such, is unique in its ability to give a rapid overview of the design and analysis of animal experiments, says Karp. “There isn’t anything else quite like this system.”

Percie du Sert hopes that a visual representation of experiments could become common practice, used in research papers or lab meeting presentations. Eventually, the EDA might even produce time-stamped versions to prove that an experiment was conducted and analysed as designed, she adds, rather than being the product of a scientist searching for meaning in data after the fact — a frowned-upon practice sometimes called HARKing (‘hypothesizing after the results are known’).

The online tool can seem a little complicated, says Jeffrey Mogil, who studies pain at McGill University in Montreal, Canada. “But I actually think that people might get a big kick out of using this,” he says. “It looks like a cool way to break in new grad students or teach the scientific method to undergrads.” ■

**“We want researchers to think about these issues at the design stage.”**

# CAREERS

**SAFETY CATCH** A gun researcher turned to an unusual funding source **p.131**

**SALES TIPS** Marketing techniques to advance your career [go.nature.com/hkpdhx](http://go.nature.com/hkpdhx)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

YASKII



LAB LIFE

## Lone-parent scientist

*Limited institutional resources mean that single parents often need a network of support to further their scientific careers.*

BY HELEN SHEN

On the Saturdays when immunologist Patricia Castillo has an experiment on the boil, she hops into her car and heads to the laboratory at 6 a.m. A postdoctoral researcher and a single mother, she aims to return home before her two sons — one aged 13, the other aged 9 — notice that she has gone.

Castillo and the father of her children separated in 2010, while she was in her third year of graduate school at the University of California, Davis. Both of her boys needed before- and after-school day care back then, and the expense ate up two-thirds of her student stipend. She received multiple childcare grants that were available to student parents at the university, and took out student loans that reached a total of US\$80,000. But after she finished graduate school, it got worse for Castillo: as a postdoc, she no longer qualified for the grants or student loans. She had to withdraw her younger son from before-school day care

and now must juggle her schedule to make things work. In practice, this means that she arrives later at the lab in the mornings and spends more of her weekends catching up on tasks. “Financially, it’s still kind of stressful. At the end of the month, I’m just barely making it till the next pay period,” says Castillo.

Castillo’s experiences as a single parent are not unusual. Scientists of all stripes, whether they are graduate students and postdocs like Castillo or senior researchers in academia and industry, face a shortage of dedicated resources to help them to balance the demands of both their career and family. Although some receive help from the ex-partner with whom they share custody, many will go it alone.

When childcare expenses or demands pile up, experiments run over or children fall ill, scientists might also turn to family, friends or hired helpers to get it all done. Many, like Castillo, spend more on childcare and other help than they can comfortably afford. And because research fellowships, programmes and policies that specifically support single

parents are rare, most scientists must look for their own solutions.

As they improvise, some have found unexpected support — and have even negotiated flexible working — by opening up to their supervisors about the challenges that they face. A number of universities offer limited childcare subsidies to postdocs and employees, as well as their students, through institutional benefit programmes. But there is no one-size-fits-all solution, and single-parent scientists must find creative ways to meet the needs of their family.

### CREATIVE PARENTING

It took a network of helpers, a lot of money and a sympathetic senior colleague to help astrophysicist Sara Seager to keep her head above water after she became a single parent in 2011. At the time, she also had a new job to learn — fast. A professor at the Massachusetts Institute of Technology (MIT) in Cambridge, Seager’s first husband died of cancer when their sons were 6 and 8 years old. Until then, her husband had been ►



► the one to handle grocery shopping, cooking and other household tasks while she focused on her single-minded search for Earth-like planets.

Emotionally and physically exhausted, Seager invited a friend to live with her in exchange for getting her boys ready in the morning and taking them to school. She also hired a babysitter to pick up and stay with the children after school. And she employed a housekeeper to prepare food and to clean the house. “I spent more than I earned, and I had to plough through my savings,” she says.

When the dean of the school of science at MIT enquired about how she was doing one day in 2012, Seager admitted that she was struggling. To her surprise, he asked what she would need financially to make things work. MIT then offered her a salary supplement that will last until her youngest child turns 16. (Seager subsequently won a no-strings, \$625,000 MacArthur fellowship and has remarried — further helping her to cope with the demands of family life.)

“As a single mom, you really have to streamline your life more than most other people do. Delegating and spending money where you can just helps so much,” says Seager. “You can’t do it all alone.” And as she discovered, existing networks of friends and neighbours can also be a valuable source of emotional and practical support.

### LOCATION CHANGE

For biologist Florence Roulet, sharing the custody of her children and openly communicating with her employer have been crucial for achieving a positive work–life balance. She and her partner, also a scientist, split in 2008, but the pair have continued to live near each other — even moving from Canada to the United States and back — to share the parenting of their twin sons.

At the time of the break-up, Roulet was preparing to leave Canada for a postdoctoral fellowship at the US National Institutes of

Health (NIH) in Bethesda, Maryland. With two four-year-old children to care for, “it was a tough time,” she recalls. But in an unconventional move, Roulet’s former partner agreed to relocate, and he also secured a job at the NIH. They began to alternate the weeks in which each looked after the children — a schedule that continues to this day.

Roulet was upfront about her family responsibilities before starting her job at a biotechnology company in Burlington, Canada, where she coordinates clinical trials and oversees regulatory applications. Although her work duties mostly fit into a nine-to-five schedule, Roulet does need some flexibility when she is looking after the children. “I had a very clear discussion with the person hiring me that I had children: ‘This is my reality. They will be sick, and I will have to go home early sometimes,’” she says.

She packs meetings, work functions and personal plans into the weeks that her ex-partner has their children. When they are living with her, Roulet sometimes works from home or at night — with the approval of her manager, who supports the flexible working arrangement as long as she continues to meet her goals and deadlines. “There are times when I need to work all night to finish up something, and I will do that,” she says.

Research fellowships and grants that are aimed at single parents are scarce, although some programmes do exist to help parents to attend conferences for short periods of time (see ‘Travel assistance for scientists with families’). And longer-term assistance is available through a small number of programmes that are designed to help scientists with significant carer responsibilities.

Lily Asquith, a particle physicist in Brighton, UK, secured one such award at a crucial moment. When her daughter Jessie suddenly became ill, Asquith was forced to consider whether to leave the research career she had worked long and hard to establish.

A single parent from the age of 19, Asquith was living in a women’s shelter when she decided to take evening classes in mathematics and physics. She was then accepted into an undergraduate degree programme at University College London. Her low income meant that she could put her daughter in day care for a reduced fee, and after Asquith received her PhD in 2010, the pair moved to the United States so that Asquith could pursue postdoctoral research at Argonne National Laboratory in Lemont, Illinois.

In 2012, when Asquith’s work took her to CERN, the European laboratory of particle physics near Geneva, Switzerland, her now teenage daughter decided to stay with her aunt in the south of England. So when Jessie was diagnosed with ulcerative colitis in 2013, Asquith was frantic to return to the United Kingdom, even without the prospect of continuing her research.

## GROUNDING WITH CHILDREN

### *Travel assistance for scientists with families*

Attendance at conferences and other gatherings of scientists is often an important part of the research process, but it can also be one of the most difficult aspects for scientists who are single parents. “Most people have no concept of how hard travel is for a single parent,” says Sara Seager, an astrophysicist at the Massachusetts Institute of Technology in Cambridge. For two years after she was widowed, she cut back on going to conferences and invited seminars. The cost of hiring round-the-clock carers to stay with her children, as well as the emotional stress of worrying about her family from afar, outweighed the professional benefits. But Seager notes that she was far enough into her career to not worry about jeopardizing her professional advancement. “I could just say no to things.”

Scientists who are at an earlier stage of their career face a tougher choice. They often feel that their advancement is tied to an ability to follow research opportunities wherever they arise, or to travelling to conferences where they can share their findings with more-senior scientists.

To help parents to attend research conferences, some universities have established grant programmes that partially offset the associated costs of childcare. For example, postdoctoral researchers and assistant professors from the David Geffen School of Medicine at the University of California, Los Angeles, can apply for a travel award of up to US\$500 to cover childcare, travel and registration for professional meetings. And Yale University

in Connecticut offers reimbursements of up to \$1,000 per year for childcare expenses that relate to travel for postdocs and certain faculty members. Similar programmes exist at institutions such as Brown University in Rhode Island, Harvard University in Massachusetts, Northwestern University in Illinois and Stanford University in California.

A number of scientific societies also offer support for parents who wish to attend their annual conferences. This year, for example, the Entomological Society of America in Annapolis, Maryland, will offer on-site childcare for children who are aged 2 or older — at no cost. (Usually, the society offers grants of up to \$400 toward childcare expenses during its conferences.)

And the US Society for Neuroscience and the American Geophysical Union, both in Washington DC, provide childcare programmes and services for a fee at their annual meetings. Other organizations offer assistance with childcare costs. The American Society for Biochemistry and Molecular Biology in Rockville, Maryland, provides grants of up to \$1,000, and the UK Microbiology Society offers grants of up to £1,000 (\$1,400), to help parents with costs that are associated with attending one society meeting per year. Many such awards do not cover children’s airfare or other travel costs.

To learn more about options for travel support and to verify eligibility for specific childcare grants and services, single-parent scientists should contact their parent institution or scientific society. [H.S.](#)

But in 2014, just as she was about to accept a job at a data-science company, she learned that her application for a Royal Society Dorothy Hodgkin Fellowship had been accepted. The five-year research award helps scientists with considerable carer responsibilities or health issues to pursue flexible working arrangements. The fellowship enabled Asquith to move her research to the University of Sussex in Brighton. "It was a real life-changer," she says. Now, she can stay on top of her research as well as spend time with Jessie, who has been in remission for the past few months.

Although Asquith has been able to continue her work without interruption, other scientists who are single parents might need to take a break of up to several years to tend to their families. For those researchers, the Daphne Jackson Trust in Surrey, UK, offers a fellowship that helps scientists to update their skills and return to research after a break. The NIH and the US National Science Foundation also offer options that enable scientists who take a leave of absence to extend the funding period of grants.

### EXTENDED FAMILY

The demands of work and childcare can be all-consuming for a single-parent scientist. But taking care of their own emotional needs should be a priority, too. "Social support is really important for single parents," says Seager. "You need other single parents. You need to find your demographic."

For Seager, that clan was an informal support network for widows in the town where she lives. The women met regularly for coffee and commiserated while trading parenting advice and offering each other emotional support. Seager also found support from within the lab. Her research group rallied round after her husband's death and became a sort of extended family. Often they would go on holiday with Seager and her children as an extension to conference trips. Back home, the group would venture out on weekend hikes. "The lab played a huge, amazing role," Seager says. "Ultimately, it's really about finding a social network. If you don't have family to rely on, it's the friends who can step in and take care of your kids and provide another kind of support."

Scientists who are single parents say that although the sacrifices and struggles can be arduous, the rewards are worthwhile. And the fulfilment that stems from maintaining a research career in difficult circumstances can help scientists to become more effective parents. "I wouldn't have done all this," says Asquith, "if it hadn't been for the ambition to be the kind of parent I wanted for my daughter." ■

**Helen Shen** is a freelance writer in Sunnyvale, California.

# TURNING POINT

## Gun-crime analyst

*Garen Wintemute has spent his career — and more than US\$1 million of his own funds — studying firearm violence. The physician-scientist at the University of California, Davis, reports that a new generation of gun-violence researchers is emerging as funding picks up.*

### Were funds available when you began this work?

Yes. In the late 1980s, rates of firearm violence were rising rapidly, and Congress made research funds available to attract people to the field. But in 1996, that mobilization effort was choked off. It was never an outright ban on research: then-US Representative Jay Dickey (Republican, Arkansas) introduced an amendment stating that funds from the Centers for Disease Control and Prevention could not be used to advocate for or promote gun control. But the writing on the wall was 'don't fund the research'. That was applied to budget appropriations, including for the US National Institutes of Health (NIH). My group and others lost funding.

### How were you able to continue your research?

We had to let people go, but we secured funding from the National Institute of Justice and private foundations. It wasn't enough.

### Did that prompt you to use your own funds?

I started to spend my own money in 2005 because I wanted to bring people together and keep this work going. Some of it can only be done in California, because we collect high-quality individual-level health and criminal-justice data related to firearm violence. In 2014, I wrote a pledge to give more over the next 4 years, up to \$2 million.

### What were your key findings?

One project established that, for people who buy guns legally, previous convictions for violent misdemeanours confer great risk for future violence. We also did the first prospective study tracking handgun purchasers and their incidence of violence: in the first week of gun ownership, the risk of firearm suicide is 57 times higher than expected for adults in California.

### Has the funding situation changed?

Yes. Last year, Dickey said he has regrets — he meant for the amendment to cut off advocacy, not research. On 11 February, he expressed support for California legislation to establish a Firearm Violence Research Center at the University of California. The person who had the most to do with funding being cut off is in a uniquely influential position to advocate for its increase. As it stands now, the NIH is funding research.



### What is the current status of the field?

I used to worry about who would do this research after I retired. There were maybe 12 of us around the country, all of a similar age. Without funding, there was too much uncertainty for most people to enter the field. But I don't worry about that anymore. We're now hiring three nationally ranked junior faculty members to join us as investigators, and launching a fellowship programme.

### What are your conversations with early-career researchers like?

The field is controversial and can be physically risky. We get hate mail and death threats. But there's plenty of intellectual elbow room and hugely important questions nobody is answering. Since the 2012 Sandy Hook shooting in Newtown, Connecticut, I have heard from all kinds of people — undergraduates to early-career professors — asking how they can help.

### How do they feel about the risks?

People have become more tolerant of the risks involved in this work. When there is something preventing research from being done, that thing feels like a bully. And no one likes a bully.

### Do you think the field will continue to grow?

Although mass shootings haven't resulted in congressional action, research funders have stepped up. I think we're still at the beginning of the beginning of a long-term change in the way the country thinks about firearm violence. We're setting up the infrastructure and labour force to keep this work going. All that said, compared to the need, the situation is still very grim. ■

### INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity. See [go.nature.com/er8c4g](http://go.nature.com/er8c4g) for more about his work.

# AJDENIA

*Let there be sunlight.*

BY NATALIA THEODORIDOU

It's cold and sunless in the tunnels. Bart is sweeping the walls vigorously, trying to warm himself up inside his stiff uniform. He scans every crack he comes across for contaminants, seals it, then moves on, sweeping and scanning, scanning and sweeping. The job is repetitive; he's heard of people who've been driven mad by the repetition of it, lost their grip. But Bart doesn't mind. If only it weren't so cold. If only there were some Sun in the tunnels. But of course there isn't.

"Sun is precious, Sun is rare," he whispers behind his mouth cover. "Sun is for the worthy." *And for Ajdenia*, he adds silently.

He tries to recall the warmth of Sun on his skin. He almost succeeds.

*And the worthiest of all get to live above, showered in sunlight.*

He finishes the length of tunnel number 8 and enters the Ra intersection. All the intersections are named after long-lost deities associated with the Sun. The next one is called Helios. Then there's Tona-tiuh, Solar Logos, Surya. Bart knows them all, every name, every inch. He always works methodically. He's good at what he does. He's worth five minutes in the filter room, under the Sun. Maybe next year he'll be worth five and a half. Or six, even.

There's another employee in tunnel number 9. Bart takes a moment to observe them. Their uniforms are identical. Their mouth masks, their goggles, their hoods. He wonders what that person looks like underneath. He wonders what they are worth. Do they also spend a few moments each day trying to remember the feeling of Sun on their skin? Are they about to lose their mind?

He continues his work on the intersection, sweeping and scanning, scanning and sweeping. The next time he looks, the other person has disappeared behind a turn in their tunnel.

Bart has finished an entire section of wall when the alarms go off, ear-piercing. The ceiling lights switch to the highest setting, bright, almost blinding. Bart puts down his scanner, sealer and sweeper, and heads

towards the centre of the intersection, as he's supposed to. He's almost there when a girl comes running out of tunnel number 7 and bumps into him, nearly throwing him off balance. He grabs her arm without thinking, steadies her. Her uniform is torn. She's not wearing a mouth cover. Bart can see her eyes behind her goggles. He would expect them to look frightened, but they are not. There is something else in there. Something bright. It makes Bart think of the Sun. It makes him think of Ajdenia.

"What are you doing?" he asks. "You're not supposed to be here."

She brings a finger to her lips. "We can live under the Sun, you, me, all of us," she whispers. "They are lying." And then she lets go of him and she's off, running into tunnel number 5.

Bart thinks of following her, but he knows he's not supposed to. He's supposed to stand in the middle of the intersection and wait for the alarms to go silent, for the lights to go back to normal. So that's what he does.

Soon, a pair of guards come out of tunnel number 7, helmets shiny and batons in hand.

"Which way did she go?" one of them asks.

"Who?" Bart blurts out, and immediately receives a blow to the ribs from the guard's baton.

"Your cooperation will be rewarded," the other guard says. "Two more minutes under the filter will be added to your next payment, if the information you provide proves correct." His tone implies that something will be taken away if not, but the exact nature of it is left vague.

*Two whole minutes of light*, Bart thinks. *Two whole minutes of Sun.*

As if noticing his hesitation, the guard who struck him scans Bart's forehead, proving they'll keep their word, one way or another. "Come on," he says. "Spit it out!"

*Could it be true?*

Bart raises his hand and points towards tunnel number 5.

Bart is in the filter room, waiting for his payment. He's thought of the girl in the Ra intersection often, ever since the day he found himself in her way. He's thought about what might have happened to her, and about her words. Could they really live under the Sun, without slaving away in the tunnels in exchange for a few moments under the protection of the filters? *She was probably one of those who lost their minds in the tunnels*, he assured himself in the end, *one of those who didn't know how to deal with the cold and the repetition, how to make themselves worthier than they are*. Bart shuffles in his chair. *And if not...* but as the thought crosses his mind, the time finally comes, and the lid of the filter room opens, letting in the Sun. Bart unzips his side pocket and brings out the plastic pot with the pink flower growing in it. He raises it towards the light. "Drink up, Ajdenia," he whispers.

He watches the pink petals shine against the Sun until the lid comes on again. ■

*Natalia Theodoridou is a cultural-studies scholar and a writer of speculative fiction. Her writing has appeared in Clarkesworld, Daily SF, Escape Pod and elsewhere. Find out more at [www.natalia-theodoridou.com](http://www.natalia-theodoridou.com).*

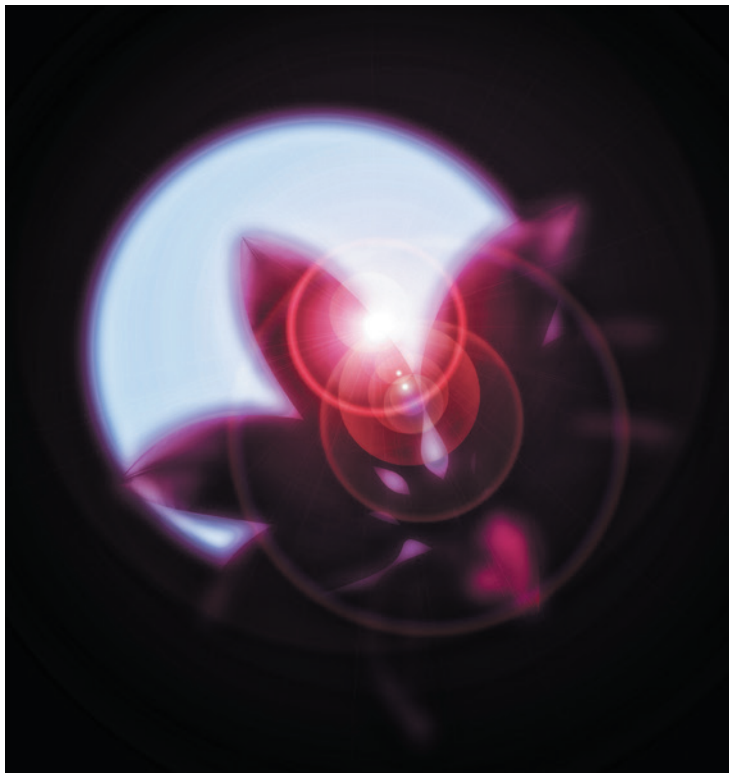


ILLUSTRATION BY JACEY



## COGNITIVE HEALTH



Produced with support from:



Research

Pushing the limits  
of brain power

# natureOUTLOOK

## COGNITIVE HEALTH

3 March 2016 / Vol 531 / Issue No 7592



Cover art: Mario Wagner

### Editorial

Herb Brody,  
Michelle Grayson,  
Richard Hodson,  
Jenny Rooke

### Art & Design

Wesley Fernandes,  
Mohamed Ashour,  
Andrea Duffy

### Production

Karl Smart,  
Ian Pope,  
Matthew Carey

### Sponsorship

Reya Silao,  
Samantha Morley

### Marketing

Hannah Phipps

### Project Manager

Anastasia Panoutsou

### Art Director

Kelly Buckheit Krause

### Publisher

Richard Hughes

### Chief Magazine Editor

Rosie Mestel

### Editor-in-Chief

Philip Campbell

From superfoods to brain training, the Internet is full of advice on how to improve cognitive health and boost brain power. Yet anyone curious enough to dip into the scientific literature will find a complicated picture behind the claims. There are seemingly contradictory studies, and gaps in our knowledge. This Outlook touches on a number of areas concerning how researchers are working to preserve or enhance this most human of faculties.

It is an unavoidable fact of life that cognition worsens as we age. For most people, this decline is imperceptible and gradual, whereas for others it is more rapid. Several approaches could prevent or even reverse the decline: common drugs, including an asthma medication (S4), and interventions to increase social interaction (S14) are both showing promise.

Some researchers go further and think that it is possible to gain cognitive benefit in youth. But claims about brain training are controversial: results vary widely and even meta-analyses come to contradictory conclusions about effectiveness (S10).

If brain training doesn't work, at least all that's lost is time. Other approaches carry real risks. Transcranial direct-current stimulation aims to activate the brain by applying a burst of electricity, and its simplicity has spawned a do-it-yourself subculture. Although the technique has therapeutic potential, it is far from clear how it works — and zapping the brain is not to be taken lightly (S6). Similarly, the rise in the use of smart drugs has many people worried (S2). But there could be one safe way to improve cognitive performance: let our technology be a benefit instead of a distraction (S9).

We are pleased to acknowledge the financial support of Nestlé Research in producing this Outlook. As always, *Nature* retains sole responsibility for all editorial content.

**Michelle Grayson**  
*Senior editor, supplements*

## CONTENTS

### S2 SMART DRUGS

#### **A dose of intelligence**

Is there any benefit to pill popping?

### S4 AGEING

#### **Restoration project**

Progress towards brain rejuvenation

### S6 NEUROSTIMULATION

#### **Bright sparks**

Why home users electrify their neurons

### S9 PERSPECTIVE

#### **Time to expand the mind**

Technology could be the best medicine, say Nicholas S. Fitz and Peter B. Reiner

### S10 BRAIN TRAINING

#### **Memory games**

Striving for better studies in the field

### S12 BRAIN FOOD

#### **Clever eating**

Do our modern diets need meat?

### S14 SOCIAL NETWORKS

#### **Better together**

Relationships keep brains healthy

### S16 NEURAL MODELLING

#### **Abstractions of the mind**

The push to keep brain models relevant

### S18 NEUROBIOLOGY

#### **Rise of resilience**

Why some people can withstand stress

## RELATED ARTICLES

### S20 Progress and challenges in probing the human brain

R. A. Poldrack & M. J. Farah

### S29 Nutrition and brain aging

P. Barberger-Gateau

### S34 From stress to social behaviour—glucocorticoids and dopaminergic circuits pave the way

J. Osório

### S35 Ageing neurons need REST

K. Whalley

### S36 Stress and the brain

R. M Sapolsky

### S39 Young blood reverses age-related impairments in cognitive function and synaptic plasticity in mice

S. A Villeda et al.

*Nature Outlooks* are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at [go.nature.com/e4dwzw](http://go.nature.com/e4dwzw)

#### CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol. XXX, No. XXXX Suppl., Sxx–Sxx (2016).

#### VISIT THE OUTLOOK ONLINE

The *Nature Outlook Cognitive Health* supplement can be found at <http://www.nature.com/nature/outlook/cognitive-health>. It features all newly commissioned content as well as a selection of relevant previously published material.

All featured articles will be freely available for 6 months.

#### SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe: Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

#### CUSTOMER SERVICES

Feedback@nature.com  
Copyright © 2016 Nature Publishing Group





Organizers of the ESL One Cologne competition tested video gamers for smart drugs for the first time last year — all the tests came back clean.

#### SMART DRUGS

# A dose of intelligence

*As mind sports becomes the new frontier for doping concerns, research is exploring whether users really get any value from 'smart drugs'.*

BY AMBER DANCE

In August 2015, 80 professional video gamers from around the world gathered for the ESL One Cologne competition in Germany. With US\$250,000 in prize money up for grabs, pressure was high, and competition organizer ESL wanted to ensure fair play. At some point during the two-day event, a random selection of players received a tap on the shoulder and were escorted to a discreet back room where a physician awaited.

For the first time in its 16-year history, ESL was taking saliva samples on its lookout for dope. Smart drugs were allegedly circulating, helping players to get in the zone. “Just like in normal sports, it’s not OK to win because you took a pill,” says Anna Rozwandowicz, ESL’s director of communications and the first head of its anti-doping initiative. That weekend, all the tests came back clean.

E-athletes aren’t the only ones allegedly popping pills to try to enhance their mental faculties. Use of the drugs seems to be common, although finding firm data is not easy. In 2014, a survey of British and Irish students reported that more than 3% currently used prescription medications as cognitive enhancers<sup>1</sup>. A 2013 survey of surgeons found that nearly 20% had used medication for cognitive enhancement

at least once<sup>2</sup>. And an informal survey ([go.nature.com/xmlrn2](http://go.nature.com/xmlrn2)) from 2008 reported that a similar proportion of *Nature* readers had used medications off-label to improve memory or concentration.

Many smart drugs are prescription medications either purchased illegally or used off-label. Top choices include Adderall (amphetamine) and Ritalin (methylphenidate) — treatments for attention-deficit hyperactivity disorder (ADHD) — and modafinil, which is a medication for sleep disorders such as narcolepsy. In people with ADHD or sleep disorders, these drugs can raise brain function so that it matches that of healthy people. But it is not clear whether the same medications can push a neurologically healthy, well-rested individual onto a higher cognitive plane. There is also the question of side effects. Despite these uncertainties, the apparently widespread use of neuroenhancers has prompted an ethical debate about whether their use is fair in school exams or mental games.

#### LIKE A BOSS

It is hard to say just how much these medications help an average person. Amphetamines improve focus and can make dull tasks seem interesting. So they might change a student’s perspective from, ‘Ugh, chemistry’, to ‘Ooh!

Carbon bonds!’ — even though that student is not any brighter. “They don’t really live up to the name smart pills,” says Martha Farah, a cognitive neuroscientist at the University of Pennsylvania in Philadelphia. “Nothing that would turn you from a B to an A student or suddenly give you winning business ideas.”

It’s still not clear precisely how these drugs produce their effects. Adderall and Ritalin are the best understood. Their main effects seem to relate to the neurotransmitters noradrenaline and dopamine, each of which mediates several effects, including attention and reward. Normally, a neuron releases these neurotransmitters as a message, telling other neurons to fire or stay quiet. Once the signal has been received, the first neuron re-absorbs the neurotransmitters. These medications block that re-uptake, so that the signals persist. Amphetamines also have other actions, such as preventing the breakdown of neurotransmitters.

Understanding of the cognitive-enhancement mechanism of modafinil is more sketchy. The drug affects “pretty much every major neurotransmitter in the brain”, says Ruairidh Battleday, a neuroscientist at the University of California, Berkeley. These include dopamine and noradrenaline, so part of its effect is probably similar to that of Adderall and Ritalin.

STEPHANIE WUNDER/ESL



Extra neurotransmitters help parts of the brain to communicate better, particularly the prefrontal cortex, which neuroscientist Kimberley Urban calls the brain's "boss". When noradrenaline and dopamine are present in the right amounts, the boss is an effective manager, explains Urban, who works at the Children's Hospital of Philadelphia. Too few neurotransmitters, and the boss is sluggish; too many, and it gets overwhelmed. The goal of treatments for ADHD and narcolepsy is to get the boss to the peak of function. Healthy users hope that they can raise their boss's peak.

People seeking a chemical boost do have legal options. "By far the most commonly used neurocognitive enhancers are nicotine and caffeine," says Peter Morgan, a psychiatrist at Yale University School of Medicine in New Haven, Connecticut. Instead of blocking reuptake, caffeine stimulates the release of extra noradrenaline and dopamine; its effects aren't as strong or as long-lasting as those provided by the other drugs. Nicotine mimics the neurotransmitter acetylcholine, which affects learning and memory. And no one bans people from pumping up their brains by smoking or drinking coffee during competitions or before an exam.

### UNCLEAR BENEFITS

Researchers are attempting to quantify the effects of prescription neuroenhancers in healthy people. In one study, Stefano Sensi, a neurologist at G. d'Annunzio University of Chieti-Pescara in Italy, and his team asked 26 university students to take an intelligence test. They then gave each volunteer either a dose of modafinil or a placebo before re-administering the test<sup>3</sup>. For the test, called Raven's matrices, participants were required to complete the ninth pattern in a 3 × 3 geometric puzzle. They were scored on the number of grids that they answered correctly. Solving the puzzle requires quick and flexible thinking — called fluid intelligence.

Results were mixed and depended on the difficulty of the matrices. Modafinil made no difference on the easiest or the hardest puzzles. The drug did increase scores for the grids of medium difficulty, mostly for those who scored low in the pre-drug test; it made little difference to participants who nailed the matrices on their first try.

Sensi's work was among 24 papers included in a 2015 review of modafinil in healthy people<sup>4</sup>. The studies used a variety of cognitive tests, and the review found that, on average, modafinil did seem to help — particularly with decision-making, planning and fluid intelligence. The more complex the task, the more that modafinil helped. "On the basis of the evidence," says Battleday, "modafinil is improving people's performance." But the



Competition at the ESL tournament was intense.

results were not uniformly positive. Not every test showed benefits and, in a couple, the drug seems to have stunted creativity.

The review authors also noted that many cognitive tests had been designed to assess impairment, not enhancement. For example, people with a brain injury or dementia may struggle with a clock-drawing test, but someone with normal cognition will usually get it right — leaving no room for smart drugs to assist. Psychologists have few options to adequately measure cognition in healthy people, says review co-author Anna Katharine Brem, a neuropsychologist at the University of Oxford, UK.

As with any mind-altering drug (caffeine and nicotine included) addiction or dependence are concerns. People who take drugs for ADHD do not seem to get hooked, says James McGough, a child and adolescent psychiatrist at the University of California, Los Angeles. However, he does not know if the same drugs might prove addictive in healthy people. After all, Adderall is an amphetamine, which has established addictive properties. Ritalin and modafinil seem to be less addictive, says Urban, but that does not mean that regular use is without risk. Morgan points out that regular use of coffee and cigarettes causes consumers' brains to adapt so that they need the stimulant just to function at their normal cognitive level. He suspects, the same might occur with smart drugs, even if users lack the compulsive craving that characterizes addiction.

As for long-term effects, nobody knows. Urban and colleagues' experiments in rats indicate that Ritalin could be bad for developing brains<sup>5</sup>. The researchers treated both adults and juveniles with one milligram per kilogram body weight, which is within the normal range for human treatment. In the grown-up rats, the drug increased nerve firing in their prefrontal cortex. But in the 15-day-old rats, equivalent to a preteen human, firing went down. If Urban stopped the drug, the effects went away. But when she tripled the dosage — equivalent to a high, but not unheard of, human prescription — the firing rates stayed low even 70 days after the treatment stopped.

The neurotransmitters that the medications

are known to target, noradrenaline and dopamine, are crucial regulators of brain maturation during puberty, Urban explains. Although the drugs don't seem to cause problems in teenagers with ADHD, they might throw off development of a healthy brain. She speculates that the poor firing patterns observed in the rats might translate to problems with working memory and flexible thinking in people. For example, someone might have a hard time finding a new route to work if their usual path is blocked. Indeed, she says, healthy children who take too much Ritalin can exhibit "extreme persever-

ance" — for example, being unable to pause a video game when it's time for dinner, persisting with one topic of conversation without being able to switch gears, or feeling emotions such as anger for a longer time than a situation warrants.

### SMART MORAL COMPASS

Smart drugs are still primitive, Sensi says. They temporarily alter multiple neurotransmitters, so they aren't very specific. A better approach, he suggests, could be to develop drugs that would promote nerve-cell growth or the rewiring of the brain, inducing changes that would permanently enhance thinking.

However, the current medications are still potent enough to raise ethical questions. One such concern revolves around social equality. Not everyone has equal access to smart drugs, and there is a danger that only the privileged will be able to get ahead with amped-up cognitive powers. The result would be yet another force widening the gap between the haves and the have-nots, says Nita Farahany, a bioethicist at Duke University in Durham, North Carolina. Or, there might be a sort of arms race, with people taking ever more advanced smart drugs just to keep up. At her own institution, Farahany says, students were so worried about brain-boosters that the university amended its honour code in 2011 to state that "the unauthorized use of prescription medication to enhance academic performance" was a form of cheating.

For now, at least, even the scientists who study smart drugs aren't relying on them. Most of those interviewed for this article said that they stick to coffee, tea or energy drinks. Morgan, for his part, suggested that the same cognitive benefits can be achieved by simply taking a refreshing nap. ■

*Amber Dance is a freelance science writer based in Los Angeles, California.*

1. Singh, I., Bard, I. & Jackson, J. *PLoS ONE* **9**, e105969 (2014).
2. Franke, A. G. et al. *BMC Med.* **11**, 102 (2013).
3. Esposito, R. et al. *PLoS ONE* **8**, e69224 (2013).
4. Battleday, R. M. & Brem, A.-K. *Eur. Neuropsychopharmacol.* **25**, 1865–1881 (2015).
5. Urban, K. R., Waterhouse, B. D. & Gao, W.-J. *Biol. Psychiatry* **72**, 880–888 (2012).



## AGEING

# Restoration project

*Future generations may have less to fear from cognitive decline thanks to microscopic insights into the ageing brain, and interventions from unexpected quarters.*

BY ANNABEL MCGILVRAY

**A**geing rodents have had a good few years. In the United States, neurologist Tony Wyss-Coray made old mice think that they were young again by giving them plasma donated by students. Thousands of miles away in Austria, neuroscientist Ludwig Aigner achieved a similar feat in elderly rats using a common asthma medication.

These are preliminary results, but both researchers are attempting to rewrite the story on cognitive decline. “It’s important to think about the aged brain differently from how we used to think about it,” says Aigner, who is head of the Institute of Molecular Regenerative Medicine at Paracelsus Medical University in Salzburg. Rejuvenation might be, as he says, “a sexy term”, but it is one that researchers such as Aigner and Wyss-Coray are using seriously in relation to preventing cognitive loss in the healthy ageing brain. Ageing does not have to

be a one-way, downhill street. But reversing the direction requires a better understanding of why the brain begins to decline in the first place.

## KEEPING NUMBERS UP

In cerebral terms, ageing starts early, long before symptoms manifest. The typical human brain begins to shrink at about age 20; by the time it is 100 years old, the brain has lost 20% of its mass. And this is for a healthy brain. Those affected by neurodegenerative disorders such as Alzheimer’s disease dwindle even more.

Ageing involves the gradual deterioration of the myelin sheathing that surrounds some nerves, and which, alongside glial cells, comprises the brain’s white matter. There is some loss of neurons, but this — in a revision of the traditional dogma — is not the main driver of the decline. Instead, neurons have reduced function, and the connections between them are weakened.

Decline isn’t uniform in the brain or across

cognitive tasks. Particularly hard hit are the hippocampus, which is crucial for memory formation and short- to mid-term storage, and the prefrontal cortex, which manages decision making and planning. This means that when deterioration occurs, memory of recent events and facts fades quicker than items that have been stored for a long time, such as words and numbers.

“There are a myriad of changes compared to the young brain that cross every level, every cell type,” says Wyss-Coray, who is at Stanford University in California. In mice, he says, the changes begin at 6 months, which is equivalent to a 30-year-old human. By the time mice reach an age that corresponds to the 50s and 60s, he says, “there are already striking changes at the cellular level in the brain”. These include modifications to the way DNA is expressed, epigenetic alterations that affect which genes are active and marked shifts in the communication between cells. Furthermore, metabolism declines in both white and grey matter as the cell’s powerhouses, mitochondria, start to fail.

At Mount Sinai School of Medicine in New York, neuroscientist Mark Baxter and other researchers at the institute are exploring the impact of ageing on synapses — the connections between neurons. His colleagues used electron microscopy to examine changes in the synapses of the hippocampal and prefrontal cortex regions of ageing rhesus monkeys. The monkeys were given tasks that tested working memory, which is associated with the temporal lobe (where the hippocampus is located). The researchers used a delayed-response task in which the monkeys needed to remember one object for a period of time and then select a second non-matching object in order to receive a food reward<sup>1</sup>. The team found that monkeys that had more problems on the memory task tended to have more axon terminals (the ends of neurons which connect to other neurons) with a single or no synapse. “The underlying synaptic structure of those neurons is being degraded,” says Baxter.

At the other ends of neurons are dendrites, which connect to axon terminals of other neurons by dendritic spines. Neurons in the prefrontal cortex have three different types of spines: stubby, mushroom-shaped and thin. The latter emerge from, and retract into, the dendrites as required, giving the prefrontal cortex great plasticity. As the brain ages, synapses in this region become dominated by mushroom-shaped connections, which are associated with long-term memory formation, but not with mental flexibility (see ‘Thinning out’). Similar to the altered synapses that Baxter’s colleagues saw in the monkey’s temporal lobes, ageing monkeys with declining working memory have an altered synaptic environment in the prefrontal cortex. “The thin spines are lost,” says Baxter.

These structural changes do not rule out therapeutics to treat cognitive decline,



according to Baxter. “The basic elements of neural computation, the neurons, are still there to work with,” he says. One possible way to preserve both the thin dendritic spines in the prefrontal cortex and the multisynapse connections in the hippocampus is to use oestrogen — in women at least. In older female rhesus monkeys, treatment with oestrogen increased the density of the thin spines and correlated with enhanced performance in delayed-response tasks<sup>2</sup>. The researchers experimented with a continuous delivery system and one that mimics the menstrual cycle. So far, they have found that the latter works best in female monkeys, “if you take oestrogen, you want to take it cyclically and not continuously,” says Baxter.

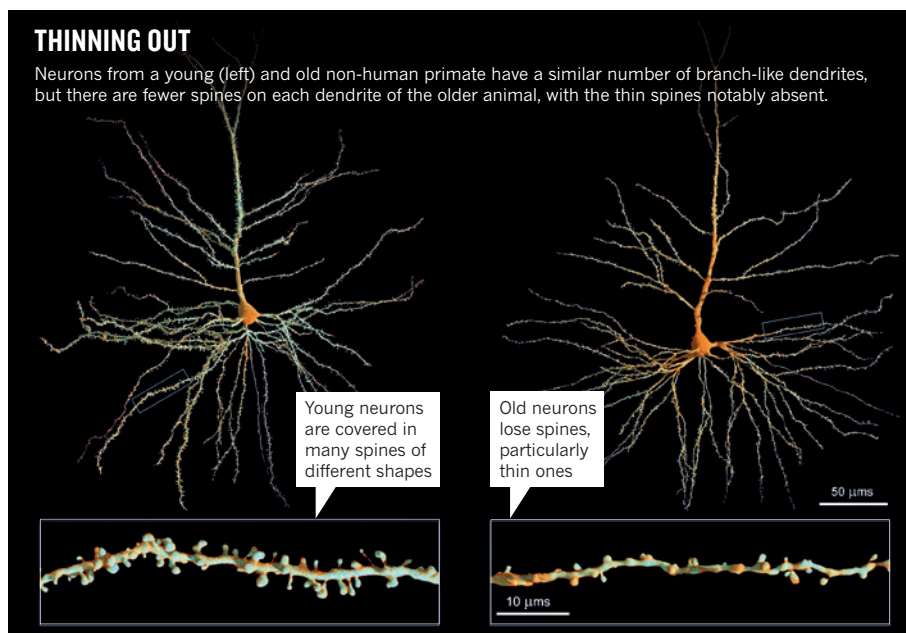
### A DOSE OF YOUTH

Wyss-Coray began his work with mice and plasma by studying the effects of parabiosis — the joining of two animals so that they share blood — on the ageing brain. He and his colleagues showed that when older mice share circulation with younger mice, the brains of the older mice show dramatic changes<sup>3</sup>. Neuron growth is boosted, and there are improvements in the animals’ ability to learn and speed of recall.

The same effect is seen when plasma from young volunteers is transfused into old mice. Once every 3 days for 3 weeks, Wyss-Coray’s team gave old mice with cognitive impairment 15 microlitres of plasma (the equivalent of one unit — around 200 ml — for humans) donated by students. The treated mice were able to remember the location of one hole among many that would allow them to shelter from a strong light shining on the testing platform. The untreated mice could only find the hole through trial and error. Is it rejuvenation? “Rejuvenation in mice, definitely,” Wyss-Coray says.

Wyss-Coray has now begun a clinical trial of the use of plasma from young donors to treat patients with Alzheimer’s disease. The double-blind trial was originally scheduled for completion in October 2015, but it is progressing slowly. There is no lack of willing participants, he says, but recruitment has been difficult owing to participation criteria that exclude certain common medical conditions and medications. Over the course of a month, half the trial participants receive weekly infusions of one unit of plasma from young (under 30 years) male donors; the remainder receive saline solution. The participants are then tested for memory, language, spatial orientation and visual attention.

In the meantime, work is continuing to determine what molecules underpin the parabiosis effect. Wyss-Coray suspects that the answer may come from outside the brain. He and his colleagues, including Aigner, have already identified eotaxin — commonly associated with asthma — as one protein that is present in higher quantities in the plasma of older



animals<sup>4</sup>. This research prompted Aigner to look at other asthma signalling proteins — and led to the discovery of the rejuvenating powers of the common asthma drug montelukast.

### INFLAMMATORY WORK

Aigner’s breakthrough attracted headlines around the world in 2015. His team showed that montelukast improved cognition and boosted neuron growth in old rats<sup>5</sup>. He stops short of saying that the drug completely restores function. “But yes, we can partially rejuvenate the brain.”

Montelukast blocks the inflammatory action that a class of proteins called leukotrienes can trigger in the lungs. These proteins also appear in the brain — in which, according to Aigner, the levels of an enzyme involved in the production of leukotrienes increases during ageing. Leukotrienes contribute to neural inflammation, cell death and the unnecessary activation of the brain’s immune cells, microglia, which can damage healthy neurons.

By giving rats with age-related cognitive decline montelukast over a six-week period, Aigner’s team stopped the inflammatory action of the leukotrienes and improved the animals’ cognitive function. Further testing in rats with dementia suggests that the drug causes a complete reversal in disease-related cognitive decline. Aigner is hoping to secure funding for a clinical trial of montelukast in patients with Parkinson’s disease.

It is no coincidence that Aigner and Wyss-Coray’s potential interventions against the cognitive decline of the healthy brain are similar to the possible treatments of neurodegenerative disorders such as Alzheimer’s and Parkinson’s. The ageing process in a healthy brain resembles the early stages of pathological deterioration in these types of dementia. Aigner points out that many of the microscopic markers of

neurodegenerative disease — synaptic decline, reduced neurogenesis, inflammation and even the notorious amyloid plaques associated with Alzheimer’s — are also present in the ageing brain of a healthy person.

Wyss-Coray agrees that current knowledge points to many microscopic similarities between pathological and non-pathological ageing of the brain. At the cellular level, “it’s very hard to discriminate normal healthy ageing from disease.” He speculates that if everybody were to live to 100 years old, most people would develop dementia. “If you look at a normal healthy control, most of those are on their way to getting a neurodegenerative disease,” he says.

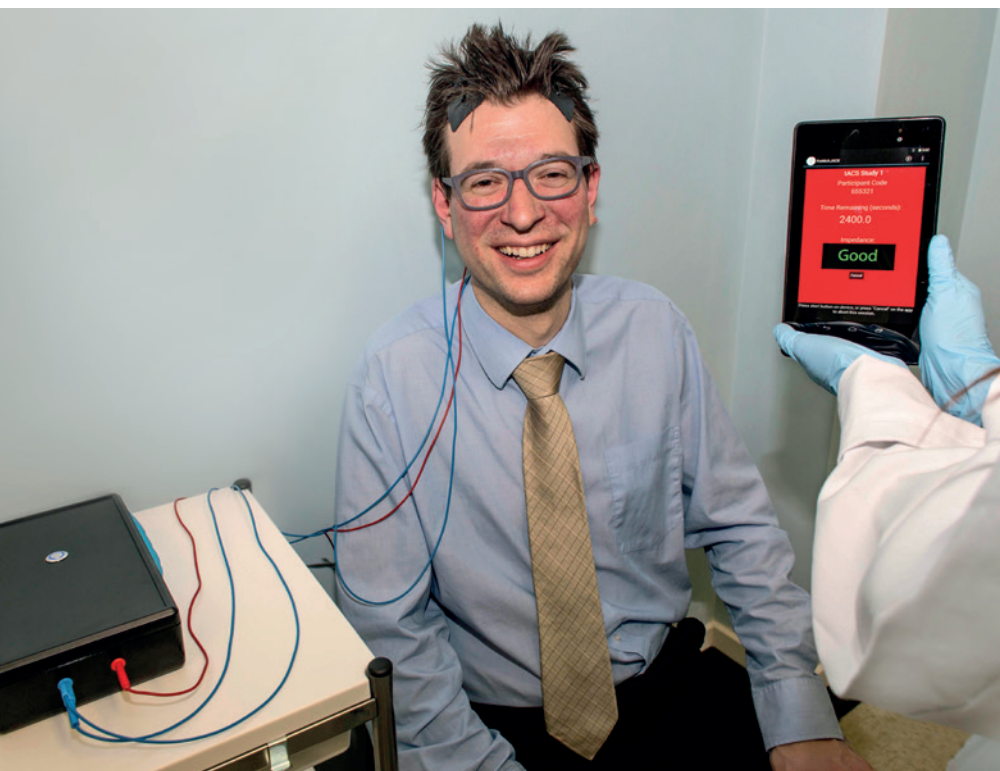
The proportion of the world’s population aged over 60 is forecast to nearly double to 22% in 2050. Cognitive impairment is already one of the leading causes of admission to residential care; without some form of prevention or rejuvenation, caring for older populations will become much more of a burden.

Neurodegenerative diseases remain the priority, says Baxter, but against a background of an ageing population, neuroscientists recognize the importance of keeping as many people as possible mentally fit. “We want everyone to have excellent cognition as they get older, so they have a good quality of life,” he says. “And they’re not struggling to remember whether they turned their stove off, or if they took their high-blood-pressure pill that morning.” ■

**Annabel McGilvray** is a freelance writer based in Sydney, Australia.

1. Hara, Y. *et al. J. Neurosci.* **31**, 7737–7744 (2011).
2. Hao, J. *et al. J. Neurosci.* **26**, 2571–2578 (2006).
3. Villeda, S. A. *et al. Nature Medicine* **20**, 659–663 (2014).
4. Villeda, S. A. *et al. Nature* **477**, 90–94 (2011).
5. Marschallinger, J. *et al. Nature Commun.* **6**, 8466 (2015).





Neurobiologist Flavio Frohlich receives transcranial direct-current stimulation.

#### NEUROSTIMULATION

# Bright sparks

*As neuroscientists explore the therapeutic prospects of brain stimulation, the amateur community are hoping the technology will enhance their mental faculties or well-being.*

BY KATHERINE BOURZAC

Lincoln walks into the neurohacker meeting I am attending in a garage in San Francisco. He takes off his flat-brimmed baseball cap, exposing two small, red burns on the side of his face. The day before, he tried a brain-stimulation method called transcranial direct-current stimulation (tDCS) for the first time. “It was pretty intense,” he says.

Lincoln had one electrode on his right temple, and another at the bottom of his left deltoid. When turned on, 2.5 milliamps of current flowed into his brain, through his medial prefrontal cortex, down to his shoulder. At least, that’s what this set-up was intended to do, he says. The idea was that the stimulation would help Lincoln, a software programmer who has meditated regularly for several years, to achieve a mystic state. That didn’t happen, although he says he did feel a heightened awareness. After 40 minutes of stimulation, he also experienced twitching in his legs.

Neuroscientists who have been studying the use of low-intensity electrical current to stimulate the brain have produced tantalizing results that have, not surprisingly, encouraged amateur use. They have shown boosts in learning, memory and performance on mathematical tests, as well as early success in treating depression and helping the recovery of those who have had a stroke. Brain stimulation is easy to do at home, either by building a tDCS set-up using some simple wiring and a battery, or buying one ready-to-use from any of the ten or so companies selling them online. Some users are seeking cognitive enhancement, whether it’s to achieve mindfulness or a memory boost; others are trying to treat mental illnesses such as depression.

If stimulation is easy, neuroscientists warn, doing it right is not. Companies selling these devices direct to consumers are “smartly circumventing government regulation” in the same way that the supplement industry is, says Flavio Frohlich, a neurobiologist at the University of North Carolina School of

Medicine in Chapel Hill. “People may well be damaging their brains.”

There is a disconnect between carefully done, quantitative research on tDCS, and more exploratory use at home or even, researchers say, in the laboratories of less-experienced scientists. Improper use and some ambiguous meta-analyses — as well as evidence of harm or negative results — have fed into a backlash against the technology within the neuroscience community. Brain-stimulation specialists are calling for a more nuanced understanding of the technology and its uses. Researchers are excited about the possibilities of brain stimulation for cognitive enhancement and therapy. They just want to take their time to validate it.

#### BIOELECTRIC BOOST

Electric stimulation has come in and out of fashion since the eighteenth century, when Italian physician Luigi Galvani famously made frog’s legs jump with an electric current, and naturalist Alexander Von Humboldt stuck wires in his back with the aim of understanding the excitability of nerves and muscles. In the nineteenth and twentieth centuries, physicians administered shock therapy to patients, inspiring fictional characters from the monster in *Frankenstein* to the horrifying Nurse Ratched in *One Flew Over the Cuckoo’s Nest*. Today’s researchers have tamed both the method and, they hope, the image of electrical stimulation.

A gentler version was popularized by research at the University of Göttingen in Germany led by neurophysiologists Walter Paulus and Michael Nitsche, who began experimenting with low levels of electrical brain stimulation in 1999. Although a typical dose of electroconvulsive therapy (which is used sparingly to treat depression) might approach 1 amp, the tDCS revived by the Göttingen group uses a tiny fraction of that — typically only 1 to 2 milliamps. That is low enough to be done with a standard 9-volt battery.

This weak stimulation cannot directly make neurons fire — instead, it generates a diffuse electrical current that changes their membrane potential. Neurons under the anode, the negative electrode through which the electrons flow, become more likely to fire when they receive signals from other neurons. Neurons under the cathode, the positive electrode, become less likely to fire. It is very difficult to target a specific region of the brain, especially with simple home set-ups that use wet sponges as the contact points.

In their early work, Paulus and Nitsche used tDCS mainly to study motor learning and working memory. But soon, many other researchers began exploring its potential for cognitive enhancement. They reported that brain stimulation acts a bit like caffeine, and may help people to learn faster. “It seems to give you any kind of benefit you want,” says Frohlich.

Such cheerful pronouncements earned tDCS

SOURCE: REF. 3

a label of ‘too good to be true.’ And indeed, the technique has now reached the backlash point in its hype cycle. Many of the positive tDCS studies have been criticized for a lack of rigour; more carefully controlled experiments are starting to show negative results. Last year, Frohlich and his colleagues wrote a report suggesting that stimulation can actually be detrimental to IQ scores. His team gave a standard IQ test to 40 people, who then received either sham or real tDCS of 2 milliamps for 20 minutes over the left or right prefrontal cortex, or both sides. When people took the IQ test again, everyone’s scores were higher (because of the well-known retest effect), but those who were stimulated actually had a smaller increase than the placebo group<sup>1</sup>. The subpar performance came in a particular part of the test that assessed fluid intelligence — the ability to solve new problems on the fly.

A few meta-analyses of tDCS studies have brought the entire field into question — and that includes lab-based research as well as the amateur use. One analysis by a group at the University of Melbourne, Australia, concluded that tDCS had “little-to-no reliable” effects<sup>2</sup>. The authors, whose analysis has been sharply criticized by the brain-stimulation community, declined to speak to *Nature* for this feature.

Nitsche and other prominent brain-stimulation specialists say that the methods used in this — and other — meta-analyses have been poor. In particular, they contend that it does not make sense to pool the results from studies that used different experimental set-ups and equipment, even if they looked at similar cognitive tasks. Marom Bikson, a bioengineer at the City University of New York, says that it would be like doing a meta-analysis of clinical trials for two drugs, only one of which works. The positive and negative results would cancel each other out, but it would be absurd to then conclude that neither drug works. What is important is not to average out the results from different electrical stimulation set-ups in a meta-analysis, but to do work that is reproducible, he says.

Brain stimulation is complicated, says Bikson. Frohlich’s IQ-deficit finding, for example, shows that there may be off-target effects that researchers miss. And the poor spatial resolution of tDCS means that researchers should design experiments carefully to make sure that they are definitely targeting the part of the brain they’re interested in, says Frohlich. Thus, even when a study leads to positive results, researchers may misinterpret the outcome unless they have carefully validated which area of the brain they are stimulating.

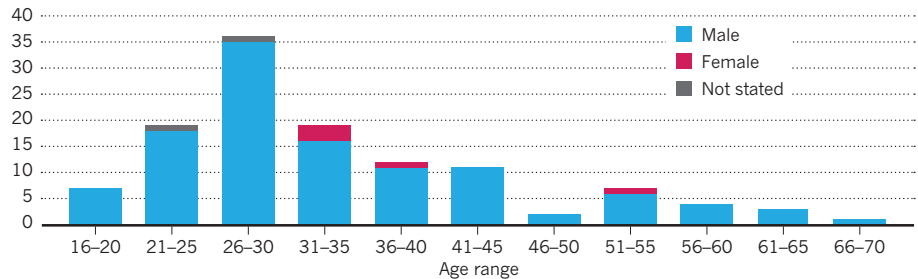
### STIMULATING MISUNDERSTANDINGS

In the scientific community, Frohlich’s work is respected by both tDCS proponents and doubters. The do-it-yourself community, however, seems to have adopted a more defensive response. One of the most popular blogs, DIY tDCS, pointed to the negative coverage under the headline “Why your

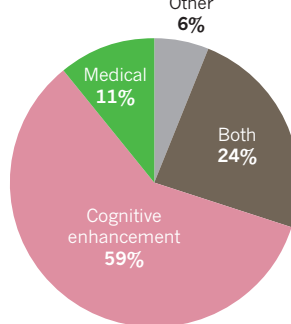
## BOYS’ OWN BRAIN BUZZ

Although the over-60s are represented, young men are the most frequent users of transcranial direct-current stimulation kits at home<sup>3</sup>. Of 121 users, most were looking for cognitive enhancement, but some were treating conditions such as attention-deficit hyperactivity disorder (ADHD) and obsessive-compulsive disorder (OCD).

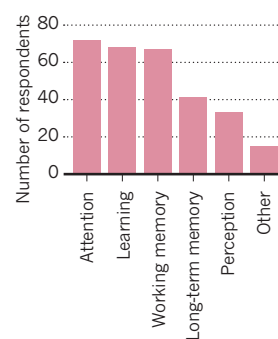
### USERS



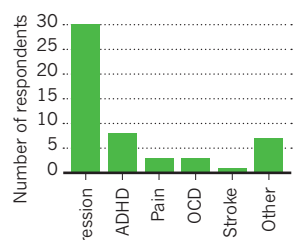
### USES



### Cognitive enhancement



### Medical conditions



brain stimulator is probably not making you stupider”, pointing out the differences between the set-ups for home users and those used in Frohlich’s study. Whether that is the case or not, unavoidable inconsistencies in the use of the home devices — from the intensity of the current to the placement of the electrodes — are troubling for researchers and brain hackers alike.

However, warnings about tDCS seem to be trickling down, at least in the San Francisco hacker community. At a weekly meet-up of the local branch of the international NeurotechX hacker organization, of which Lincoln is a member, I am chatting with six programmers gathered around folding tables and couches. They are talking about a 3D printed electroencephalogram (EEG) cap that is available online, and working on open-source software for brain-computer interfaces.

Computer-science graduate student Damien talks about the excitement of “exploring your own brain” using EEG feedback. But stimulation? No way, say most of them. “It hasn’t been proven that it’s harmless,” says software engineer Marion, who hosts the meet-up. This group takes a read-only approach, recording the brain’s electric signals, but not stimulating them. When Lincoln walks in with the burns on his face, I am not the only one to raise an eyebrow — the side effects he noticed and

the uncertainty about what really happened in his brain reinforces their scepticism. Lincoln says he wouldn’t use that particular set-up again, especially not with the same company’s tDCS device, but he might try another set-up.

It’s difficult to know what people are actually doing at home, and how many home users there are. A tDCS forum on the social-news website Reddit (known as a subreddit) had around 8,000 members at the beginning of 2016, but not all readers are necessarily users of the technology. Posts include tips on electrode placement, links to media coverage of scientific results and some alarming questions, such as one from a reader who asked whether childhood epilepsy makes tDCS more risky as an adult.

So far, Anita Jwa at Stanford Law School in California, whose research focuses on the intersection between law and neuroscience, is the only researcher who has studied home users<sup>3</sup>. Jwa says that tDCS users do not meet up in person very much, as far as she knows, and that the online community has a self-regulating aspect; the reader who asked about epilepsy, for example, was told to ask his doctor or simply not to take the risk. On the basis of surveys posted on two popular websites: the tDCS subreddit and DIY tDCS, Jwa found that users were mostly in their 20s and 30s and 94% were male.

Home tDCS users, says Jwa, tend not to see themselves as experimentalists who are adding to the pool of knowledge. “Most users are doing it for cognitive enhancement,”

*“People are desperate, they are driven to these things out of a lack of effective tools.”*





L: The burns Lincoln received from a DIY kit. R: Others have more success using a different set up.

she says (see ‘Boys’ own brain buzz’). Most people who sought cognitive enhancement wanted to boost attention, learning or working memory.

### SPEAKING THE BRAIN’S LANGUAGE

While repeating the mantra ‘don’t try this at home’, neuroscientists admit that people treating themselves for illnesses such as depression are trying to make up for the real shortcomings of mainstream medicine. “People are desperate, they are driven to these things out of a lack of effective tools,” says Adam Gazzaley, a neuroscientist and psychiatrist at the University of California, San Francisco. Gazzaley is working on combining brain stimulation with brain-training video games for cognitive enhancement.

What’s especially frustrating for neuroscientists is that brain stimulation has real promise for treating conditions and for cognitive enhancement — the very things that companies are implying their machines do (while taking pains to avoid making actual medical claims, which would trigger government regulation). Ultimately, validated devices and stimulation procedures will replace what’s available today. When researchers come up with something better, “the snake oil will go away,” says Gazzaley.

To get there, some researchers are seeking a better understanding of the mechanisms behind tDCS. “We need to find out how it works so we can make it better,” says Frohlich. Bikson disagrees — the technique is too promising to stop and wait for a full mechanism, he contends.

For those with a biophysics bent, the mystery about the mechanisms is a great motivation. Cognitive scientists Ludovica Labruna and Richard Ivry, of the University of California, Berkeley, fall into this camp. They hope to use a better-understood brain-stimulation method to try and illuminate the workings of tDCS.

Transcranial magnetic stimulation (TMS) uses a focused magnetic field to cause small numbers of neurons to fire. TMS’s direct effect is much easier to measure in humans than tDCS’s more mysterious influence. Because of this, researchers know that all kinds of things can influence a person’s sensitivity to TMS,

including skin conductivity, skull thickness and even subtle differences in brain anatomy.

Labruna shows me how TMS sensitivity can be quantified. She tapes an electrode near the web of skin between my index finger and thumb. This electrode will measure the voltage in my hand when she stimulates my brain with a TMS paddle. Labruna locates the part of my motor cortex that controls this particular muscle in the hand,

*“We need to find out how it works so we can make it better.”*

then cranks up the TMS. The machine can focus a magnetic field of about 3 tesla in a small area of the cortex; TMS sensitivity is measured by determining what percentage of this maximum magnetic strength is required to provoke a threshold potential of 1 millivolt in a muscle. After a few tries, my hand jerks like a puppet. My sensitivity is medium-high — whether it’s because my skull is thin, or something else, I respond when the field’s intensity is only 42% of its maximum (most people respond at about 50% maximum intensity; some very sensitive people respond at about 29%, others not until about 60%).

In preliminary results presented at the Society for Neuroscience meeting in October 2015, Labruna, Ivry and their colleagues showed that the more sensitive a person is to TMS, the more readily they respond to tDCS, and the better they do in a motor-learning task compared with both those who had a sham stimulation and those who were less sensitive to TMS. The researchers are now designing an experiment that will test whether this correlation holds in tDCS experiments that look at other types of cognition.

### ALTERNATE REALITY

Although tDCS has been getting most of the buzz, there are other kinds of electric brain stimulation that may work better, or at least have different applications. Some researchers, for example, are using alternating current, which targets brain oscillations, instead of direct current, on the basis of the theory that this may be a more natural way to stimulate the brain.

Transcranial alternating current stimulation

(tACS) emerged in 2006, when researchers in Germany showed that stimulating the brains of healthy people at 0.75 hertz (at the lower end of the frequency range of delta waves) during the early stages of sleep encouraged this rhythm and enhanced memory retention<sup>4</sup>. Other frequencies are associated with different cognitive states: theta (5–8 Hz) with working memory and gamma (>30 Hz) with memory maintenance, although these associations are broad and highly dependent on where in the brain the patterns are measured.

When you close your eyes and relax, the brain’s oscillations are about 10 Hz (within the alpha frequency of 7.5–12.5 Hz). In a preliminary study, Frohlich and his colleagues measured a person’s alpha frequency with an EEG, then applied an alternating current at a matching frequency. They found that this enhanced creativity<sup>5</sup>.

Like tDCS, the mechanism of tACS is not clear. One theory is that it might be more targeted because the rhythmic simulations interact with existing brain activity only at a particular frequency. The effects of tACS are also thought to be more short-term than those of tDCS, says Roi Cohen Kadosh, a cognitive neuroscientist at the University of Oxford, UK. That still needs to be proved, but it’s an exciting hypothesis, says Frohlich. Whatever the kind of current, the brain isn’t a simple machine, he cautions, and it isn’t possible to turn a function on like a light switch. “We have to speak the language of the brain and understand how the brain responds,” Frohlich says.

Despite the backlash against tDCS, many neuroscientists have no doubt that transcranial electrical stimulation will come to be an important tool for cognitive health and well-being. “The brain uses both neurotransmitters and electric fields to communicate,” says Sarah Hollingsworth Lisanby, director of the division of translational research at the National Institute of Mental Health in Bethesda, Maryland. Therefore, she says, we should use both channels for therapy. Moreover, she adds, non-invasive stimulation may be a way to intervene earlier in the development of mental illness — perhaps even to prevent it.

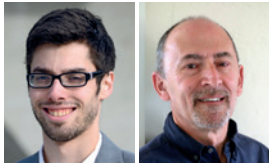
Bikson thinks in a similar way. The commonality between brain stimulation for cognitive enhancement and for therapy is that both involve learning. You are “trying to teach the brain,” he says, “either a new trick — or not to be sick.” ■

*Katherine Bourzac is a freelance journalist based in San Francisco, California.*

1. Sellers, K. K. et al. *Behav. Brain Res.* **290**, 32–44 (2015).
2. Horvath, J. C., Forte, J. D. & Carter, O. *Neuropsychologia* **66**, 213e36 (2015).
3. Jwa, A. J. *Law Biosci.* **2**, 292–335 (2015).
4. Marshall, L., Helgadottir, H., Mölle, M. & Born, J. *Nature* **444**, 610–613 (2006).
5. Lustenberger, C. M., Boyle, R., Foulser, A. A., Mellina, J. M. & Frohlich, F. **67**, 74–82 (2015).



## PERSPECTIVE



# Time to expand the mind

Thoughtful use of ubiquitous technology can improve mental ability more than drugs and devices, say **Nicholas S. Fitz** and **Peter B. Reiner**.

Few objectives are more desirable than improving mental performance. Whether it is to enhance attention and memory or to stave off the normal decline of ageing, the drive for cognitive enhancement has caught the imagination of scientists and the public. The media, in particular, regularly report strategies that offer 'limitless' abilities. Despite the hype, drugs and devices produce, at best, only modest gains<sup>1</sup>. Indeed, the current evidence indicates that the best ways to improve mental ability are the familiar approaches of proper sleep, good nutrition and exercise.

One reason may be that our brains are already working at near-optimal capacity; attempts to modify them bump up against the hard limits of neurobiology. If this is the case, then we need to find a new strategy for boosting cognitive power. The best path, we believe, will be to improve the way in which we blend our mental capabilities with the powerful algorithms in our computers and smartphones.

## PERSPECTIVE SHIFT

It is hard to overstate the degree to which information technology has permeated today's world. For many, reliance on technology begins when they awaken, continues throughout the day and ends only when they drift off to sleep. Nearly half of the adult population worldwide owns a smartphone. But to call these devices phones is a misnomer; the functions they carry out are incredibly diverse, including storing information, sending and receiving messages, and videoconferencing. Sometimes it feels as if technology is supplanting thinking, but this worry is somewhat misdirected. Rather, these devices are extending the reach of our cognitive abilities, so much so that a statement that once seemed radical is increasingly realistic: we have become proto-cyborgs<sup>2</sup>.

We are still a long way from the transhumanist fantasy of upgrading our brains with implantable computer chips. But we have entered a transitional era in which we are commingling our cognitive space with technology. In doing so, we have enlisted the assistance of what might be termed technologies of the extended mind<sup>3</sup>.

Consider the ways in which these tools already enhance cognition. Never before have humans had the ability to find answers in the blink of an eye, and to store vast amounts of data with greater fidelity than biological memory. As the Internet of Things gains momentum, we will find ourselves interacting with 'intelligent' objects that predict our preferences and make decisions on our behalf. Ideally, delegation of these tasks to our devices would allow us to expend more energy pursuing challenging activities such as improving willpower and analytical thinking.

But that is not how the human-technology connection is playing out. Instead, the same devices that extend some cognitive abilities degrade others. People grappling with information overload often bemoan the difficulty they have in concentrating for long periods

of time<sup>4</sup>. Although we seem to be constantly connected to others, some overuse technology at the expense of face-to-face social engagement. And it is the rare individual who does not succumb to the addictive appeal of devices that provide instant rewards. Turning them off is impractical because it can leave us disconnected from information that is truly valuable. Our gnawing anxiety about overusing our technology is only partly soothed by the current crop of apps, such as Freedom or Moment, that attempt to quiet the noise of digital life by blocking distracting websites for a limited period. A better strategy, however, is to embed solutions in the design of both the devices and the environments in which we use them.

## THE ATTENTION ECONOMY

One approach that merits further consideration is the idea of calm technology<sup>5</sup> — interfaces that inform without distracting.

For example, your phone might automatically go silent when it knows that you are in a meeting, unless the message is urgent. Behavioural nudges that help people to more closely align their technology use with their intentions could reinforce embedded design principles. Such alterations of the digital and physical environments can foster a healthier relationship with our devices. We endorse fledgling movements such as Time Well Spent that call for a broad coalition to work towards real-world solutions. We encourage the scientific community to investigate the implications — both beneficial and deleterious — of technology adoption. We call on policymakers and behavioural scientists to consider crea-

tive means of encouraging best practices by the public. And we challenge technology designers to improve the user experience so that cognitive health is no longer sacrificed on the altar of profit margins.

Our growing reliance on ubiquitous computing technologies defines the modern age. Melding these technologies with thoughtful design and effective nudges that reinforce healthy social norms will ensure that they genuinely improve the human condition. ■

**Nicholas S. Fitz** is a graduate student at the National Core for Neuroethics at the University of British Columbia. **Peter B. Reiner** is a professor at the National Core for Neuroethics at the University of British Columbia.  
e-mail: peter.reiner@ubc.ca

1. Farah, M. J. *Science* **350**, 379–380 (2015).
2. Clark, A. *Natural-born Cyborgs: Minds, Technologies, and the Future of Human Intelligence* (Oxford University Press, 2003).
3. Reiner, P. B. *The Neuroethics Blog* <http://go.nature.com/gl8p4z> (2015).
4. Atchley, P. & Lane, S. in *Psychology of Learning and Motivation* 133–177 (Academic, 2014).
5. Weiser, M. & Brown, J. S. *Designing Calm Technology* (Xerox PARC, 1995); available at <http://go.nature.com/wgsjlju>.

WE HAVE ENTERED A  
**TRANSITIONAL**  
ERA IN WHICH WE ARE  
COMMINGLING OUR  
**COGNITIVE**  
**SPACE**  
WITH TECHNOLOGY.



An advert for Pelmanism, a brain-training technique that became popular in the early twentieth century.

#### BRAIN TRAINING

# Memory games

*Conflicting results are expected in a young field, but what do you do when even the meta-analyses do not agree?*

BY SIMON MAKIN

**N**ew Minds for Old in 12 Weeks!" proclaimed adverts for Pelmanism, a brain-training technique that swept the United Kingdom in the early part of the twentieth century. Promotional material claimed that this system could combat such troubling mental phenomena as "lack of ideas" and "brain fog". It became so successful that the Pelman Institute established offices in Australia, South Africa and the United States. Pelmanism was still being promoted as late as the 1960s, but has since sunk into obscurity, becoming merely a curious chapter in the history of psychology.

Except that almost 100 years later, brain training is back — this time with scientific backing. The first apparent breakthrough came in 2002 when a group of researchers in Sweden showed that training children with attention-deficit hyperactivity disorder on adaptive working-memory tasks — which test a person's ability to retain and manipulate information over a short period of time

and that increase in difficulty to match performance — improved their attention and reasoning<sup>1</sup>. But the real excitement came in 2008 with research led by psychologist Susanne Jaeggi, now at the University of California, Irvine. Jaeggi's study seemed to show that healthy young adults who practised an adaptive working-memory task, which the authors called dual *n*-back (see 'A test of sight and sound'), showed increases in the unrelated ability of fluid intelligence (the ability to reason in novel situations)<sup>2</sup>. Furthermore, there was a dose effect: the more that people trained, the 'smarter' they became.

And, as with Pelmanism, a lucrative industry has grown up around the idea that cognitive performance can be enhanced by training. But the claim that mere hours spent playing a memory game can increase intelligence is an extraordinary one, and sceptics soon started voicing objections. Negative studies appeared, and the field is now awash with conflicting results. When faced with a large but uncertain evidence

base, researchers usually turn to meta-analysis to assess the evidence. Unfortunately, in this field, even meta-analyses are producing divergent conclusions. Some researchers have suggested that the inconsistencies stem from the use of inadequate control groups and measures of outcome. But most agree that the field needs bigger, better studies — and a return to basic science.

#### THE TROUBLE WITH TRAINING

Numerous studies purport to show benefits from cognitive training, but delve deeper and there is less than meets the eye. Many of these studies show little more than improved performance on tasks closely related to those that participants trained for — known as near transfer. "When you practise something, of course you get better at it," says Jaeggi. "The real question is whether there is far transfer", referring to when training benefits different cognitive abilities. This is why a lot of training research has focused on working memory. Working-memory capacity predicts everything from reading ability to academic achievement, and correlates highly with fluid intelligence. Increasing this capacity might, therefore, have a broad impact on cognition.

Among the most prolific early sceptics of far-transfer effects were Randall Engle and his fellow psychologists at the Georgia Institute of Technology in Atlanta. They pointed to two main recurring problems in working-memory training studies, the first being inadequate control groups. Many findings, they said, could be due to 'Hawthorne effects', referring to the fact that people change their behaviour when they know that they are being observed. To discount this effect, Engle and colleagues recommended that studies should use active control groups in which participants perform activities that are identical to the training in every aspect except the main task. For example, although Jaeggi and her team<sup>2</sup> tested their control group before and after training, they did not give them additional exercises; they accounted for 'test-retest' effects (people do better the second time around), but not phenomena such as the Hawthorne effect.

The second issue concerns the use of only one measure of an outcome. Participants can develop strategies during training that aid their performance on a task — for example, inner vocalization of a visual stimulus — without any change in their underlying ability. And, because no task taxes only one cognitive ability, practising one exercise can lead to improvements on seemingly unrelated tasks if there is overlap between the abilities that they engage. To avoid this possibility, the consensus among cognitive psychologists is that researchers should use a range of tasks that cover, for example, numeric, verbal and visuospatial abilities. Most studies have not done this.

There have now been several attempts to clarify the picture by pooling study results. In 2013, psychologists Monica Melby-Lervåg of Oslo

THE GRANGER COLLECTION/TOPFOTO

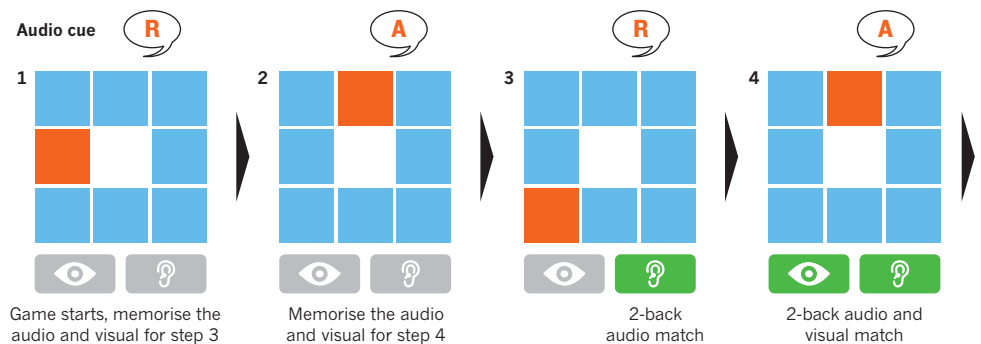
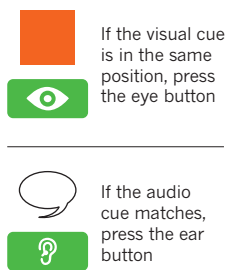
➔ NATURE.COM

To watch an animation of the dual *n*-back test visit: [go.nature.com/dofbdm](http://go.nature.com/dofbdm)



## A TEST OF SIGHT AND SOUND

In the dual  $n$ -back test, participants are presented with an audio and a visual cue simultaneously. Participants must remember both and determine whether they match  $n$  turns later. In this example,  $n = 2$ .



University and Charles Hulme at University College London produced a meta-analysis<sup>3</sup> that included 23 studies of adaptive working-memory training, each lasting at least 2 weeks. They found a small far-transfer effect on non-verbal reasoning — but only in studies that used passive control groups. In another meta-analysis<sup>4</sup> of 20 studies that only looked at  $n$ -back training, psychologist Jacky Au of the University of California, Irvine, Jaeggi and their colleagues found a statistically significant positive effect on fluid intelligence — albeit small. “There is very good evidence that training on working-memory tasks does improve performance on tests of fluid intelligence,” says Au. But, he adds, it is not yet clear whether this translates to a meaningful, real-world increase in intelligence.

The analyses did little to settle the matter and the two groups have since exchanged further critiques<sup>5,6</sup>. Hulme and Melby-Lervåg have gone so far as to call for journals to stop publishing studies that use passive controls<sup>5</sup>. “No analysis can get over having a passive control group,” says Hulme. “The only way round it is to have a control group that’s doing something very similar, with the same expectations that it’s beneficial.” But training studies are time-consuming and expensive; doubly so when using active controls. “Best practice is to use active controls,” Au admits, “but our work suggests the cost–benefit profile is not always clear, especially given the current state of funding.”

Meta-analysis is a powerful tool for finding a consensus in a confusing picture. However, it cannot make up for the design flaws of the constituent studies, says psychologist Claudia von Bastian of the University of Colorado, Boulder. And many brain-training studies have weaknesses. For example, the average number of participants per training group in Au’s meta-analysis was 20, meaning that most studies did not have enough participants for their results to be reliable — particularly when it comes to detecting small effects. This low statistical power not only increases the risk of finding spurious effects, but it also tends to inflate the size of any effect found.

Neither can meta-analysis adjust for studies that use a single (and therefore inadequate) measure of an outcome. Using multiple measures can enhance the validity of results, but some researchers advocate going further.

“There are statistical methods that allow you to analyse change in cognitive abilities instead of changes in test scores,” says cognitive neuroscientist Ulman Lindenberger of the Max Planck Institute for Human Development in Berlin. These ‘latent variables’ are inferred from the analysis of a battery of observed measures, and represent shared variance in those measures — in other words, changes in some underlying ability, such as reasoning or memory. Psychologist Florian Schmiedek of the German Institute for International Educational Research in Frankfurt, and his co-authors found that less than one-quarter of brain-training studies used multiple outcome measures, with only 7% using latent variables<sup>7</sup>. Schmiedek urges researchers to use more robust methods. “I don’t think this case will be closed by just running more of the kinds of studies that went into the recent meta-analyses,” he says. “The few studies that went to such effort paint a modest picture of the effectiveness of cognitive training.”

### BACK TO BASICS

What is clear is that the effect of working-memory training on fluid intelligence lies somewhere between zero and very small, depending on whose analysis you trust. Taking the optimistic view, the question becomes one of opportunity cost: how much time and effort is needed to produce a lasting effect, and how does this compare to other uses of that time? Other activities purported to improve cognition include exercise, musical training and learning a new language. Unfortunately, studies of these activities often have similar problems to brain training, says von Bastian, from choosing valid controls to difficulties in untangling causality — and results are similarly contested.

Other researchers are examining the problem from a biological perspective. Proponents of brain training are fond of pointing out that the brain remains plastic throughout life. But, a degree of stability is also essential. Lindenberger and cognitive neuroscientist Martin Lövdén of the Karolinska Institute’s Aging Research Center in Stockholm, argue that the brain strikes a balance between plasticity and stability, and that this shifts as we age. From this perspective, changing something as fundamental as intelligence in an adult is a big deal. “Everything from a theoretical perspective suggests that to move

intelligence you would need to do something massive,” says Lövdén.

Some studies have looked for evidence of such changes at a neural level. The problem is that neuroscientists do not yet fully understand how everyday experience affects the brain. During learning, both activity and volume in parts of the brain initially increase, but then decrease — and the time course is not clear, says Lövdén. “We don’t have the link yet between experience, the brain and behavioural change,” he says. “We need to take a step back and try to understand the basic science.”

Von Bastian also advocates a return to the fundamentals. She is focused on developing a greater understanding of working memory — what its components are, the extent to which each might be malleable and how these components might affect other cognitive abilities. “That might be very interesting as a way to experimentally look at the relationship between working memory and intelligence,” she says. Research using this kind of theory-driven design, however, has been lacking.

To facilitate the sharing of training tasks, protocols and data, von Bastian has developed a web-based, open-source software package, Tatool. The aim is to stimulate “methodologically rigorous research” with huge sample sizes that is unbiased by commercial products, she says. So far, more than 100 researchers are signed up, which, she says, should bring some new ideas and methodology to help explore the questions around whether brain training really works. “If we keep doing bad studies with small samples, we’ll never know.” ■

**Simon Makin** is a freelance science writer based in London.

1. Klingberg, T., Forssberg, H. & Westerberg, H. *J. Clin. Exp. Neuropsychol.* **24**, 781–791 (2002).
2. Jaeggi, S. M., Buschkuhl, M., Jonides, J. & Perrig, W. J. *Proc. Natl Acad. Sci. USA* **105**, 6829–6833 (2008).
3. Melby-Lervåg, M. & Hulme, C. *Dev. Psychol.* **49**, 270–291 (2013).
4. Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M. & Jaeggi, S. M. *Psychon. Bull. Rev.* **22**, 366–377 (2015).
5. Melby-Lervåg, M. & Hulme, C. *Psychon. Bull. Rev.* <http://doi.org/bb9d> (2015).
6. Au, J., Buschkuhl, M., Duncan, G. J. & Jaeggi, S. M. *Psychon. Bull. Rev.* <http://doi.org/bb9f> (2015).
7. Noack, H., Lövdén, M. & Schmiedek, F. *Psychol. Res.* **78**, 773–789 (2014).





Early humans who hunted animals for meat developed bigger brains than plant eaters.

#### BRAIN FOOD

# Clever eating

*Consumption of animals helped hominins to grow bigger brains. But in a world rich with food, how necessary is meat?*

BY SUJATA GUPTA

Around 6 million years ago, primates started moving from tropical forests into the savannahs. Unlike today, these prehistoric expanses were humid and probably provided a year-round supply of fruit and vegetables. But then, some 3 million years ago, the climate changed and the savannahs — along with their plentiful food supply — dried up.

Many mammals, including some primates, went extinct, but others adapted. Archaeologists working at sites in modern Ethiopia have discovered animal remains that date back almost 2.6 million years. The telltale cut marks on their bones are almost certainly signs of butchery<sup>1</sup>, says Manuel Domínguez-Rodrigo, a palaeoanthropologist at Complutense University in Madrid.

Only two types of primate survived the climate catastrophe, says Domínguez-Rodrigo. There was a “plant-processing machine on the one hand and a meat-eating machine on the other hand”, he says. “The meat-eating machine evolved a bigger brain.”

The meat-eating machine became us.

To build and maintain a more complex brain, our ancestors used ingredients found primarily in meat, including iron, zinc, vitamin B12 and fatty acids. Although plants contain many of the same nutrients, they occur in lower quantities and often in a form that humans cannot readily use. For instance, red meat is rich in iron derived from haemoglobin, which is more easily absorbed than the non-haem form found in beans and leafy greens. Furthermore, compounds known as phytates bind to the iron in plants and block its availability to the body. As a result, meat is a much richer dietary source of iron than any plant food (see ‘Meat efficiency’). “You would need to eat a massive amount of spinach to equal a steak,” says Christopher Golden, an ecologist and epidemiologist at Harvard University in Cambridge, Massachusetts.

The implications for cognitive health are huge. There is a clear, but underappreciated link between meat and the mind, says Charlotte Neumann, a paediatrician at the University of California, Los Angeles, who has studied meat eating in Africa and India for the past three decades. Deficiencies in the micronutrients found in meat have been linked with brain-related disorders, including

low IQ, autism, depression and dementia. Iron is crucial for the growth and branching of neurons while in the womb; zinc is found in high concentrations in the hippocampus, a crucial region for learning and memory; vitamin B12 maintains the sheaths that protect nerves; and omega-3 fatty acids such as docosahexaenoic acid (DHA) help to keep neurons alive and to regulate inflammation.

#### MEAT FOR THE POOR

In the 1980s, researchers began to suspect that a lack of meat in some poor rural villages was contributing to a spectrum of childhood problems, including short stature, weakened immunity, social difficulties and poor school performance. When researchers from five universities studied the effects of chronic malnourishment in Mexico, Kenya and Egypt, they found that children who consumed the greatest amount of meat and dairy products scored highest on physical, cognitive and behavioural tests, particularly in Kenya<sup>2</sup>. But was the absence of meat really to blame? What the researchers needed was a controlled study.

So Neumann began a trial in Kenya<sup>3</sup>. Her team selected 12 schools with children aged 6 to 14, and gave some of the children mid-morning snacks. Schools were divided into four groups: the control group was not given a snack, whereas the other three received variations on githeri, a traditional porridge that consists of maize (corn), beans and greens. One group received a basic version, the second received the basic githeri with a glass of milk, and the third had meat added; all githeri were balanced to contain the same amount of calories. The study continued for more than 2 years and spanned 2 cohorts, the first with 525 students and the second with 375. The students’ physical health and classroom performance were measured every three or six months. Compared with the other groups, students in the meat group had greater muscle mass and fewer health problems, and even showed greater leadership in the playground. Cognitive performance was stronger, too: the meat group outperformed other groups in maths and language subjects<sup>4</sup>.

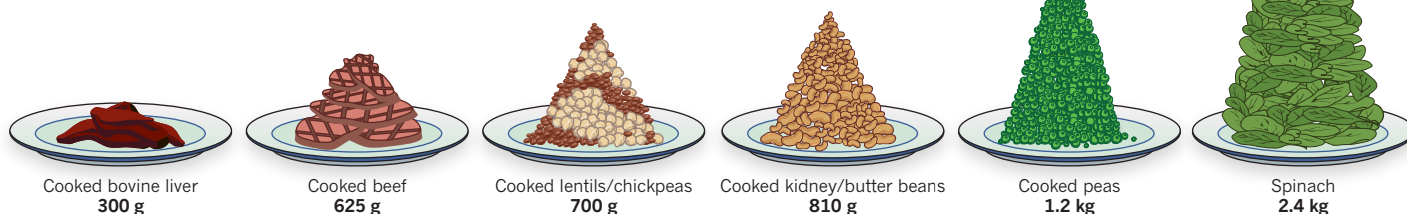
Neumann was not surprised by the results. The typical diet in rural Kenya is subsistence-based and does not include many nutrients that help the brain to grow. The challenge now is to get people to consume more meat, which is widely regarded as too expensive. What people don’t realize, Neumann says, is that to nourish the brain, pretty much any animal matter will do: “Meat can be a worm, caterpillar or termite. It doesn’t have to be butcher meat.”

#### MEAT FOR THE RICH

But how does meat fit into a richer diet? “A lot of the studies that have demonstrated the importance of meat, vitamin B, animal products and protein generally have been carried out in populations receiving very little nutrition,” says Diane Hosking, a healthy-ageing

## MEAT EFFICIENCY

To reach the recommended daily intake of 18 milligrams of iron, a woman would have to eat at least 8 times more spinach than cooked liver. Iron found in vegetables is also harder for the body to absorb, because it is usually bound to fibre.



These data are approximate and will vary depending on factors such as preparation technique, soil or feeding conditions, and time between harvesting and intake. Analysis by F. Mori Sarti based on data from <http://ndb.nal.usda.gov> and <http://www.unicamp.br/>

researcher at the Australian National University in Canberra.

To fill this gap, Hosking and her team asked 352 Australians aged between 65 and 90 years old — who were cognitively healthy and predominantly from a middle- or high-income background — to recall what sorts of food they ate growing up<sup>5,6</sup>. For instance, how often did they eat items such as carrots, meat, fish or cake? The researchers then administered cognitive tests.

Hosking found no correlation between the volunteers' test performance and their consumption of meat as children. The results contradicted what Neumann and others have observed in developing countries. What's more, contrary to conventional wisdom, participants who consumed more fish during childhood and as adults were actually slower on measures of cognitive speed. (The fish might have contained neurocontaminants such as mercury, she says.)

There are several issues that affect these results, says Hosking. One is that people don't eat single foods, but patterns of foods, making it difficult to tease out the importance of an individual food type, such as meat. In the older Australians for instance, those who ate meat were also more likely to consume packaged desserts and snack foods.

Moreover, what the animal eats also matters. Livestock and poultry in Western nations are often raised in large facilities and fed diets that consist mainly of maize and soya, whereas animals from poor villages are typically farmed on a much smaller scale and forage for a greater variety of foods, which increases the nutrient content of their meat. Given these sorts of variations, Hosking says, "we have to be very cautious about making dietary recommendations ... for people who have access to large quantities of food."

## MEAT IN THE BRAIN

The micronutrients in meat have become an essential part of our diet over millennia. A few years ago, archaeologists in Tanzania unearthed fragments of a child's skull dating back 1.5 million years. Deformities on the bones suggested that the child had died from porotic hyperostosis, a condition thought to result from a

deficiency in vitamin B12 — found exclusively in animal-derived foods. Humans started eating dairy products only in the past 5,000 years, meaning that the child had almost certainly died from a lack of meat<sup>7</sup>. So, by at least 1.5 million years ago, says Domínguez-Rodrigo, humans had become so adapted to eating meat that without it they would die.

Research is starting to provide some clues as to how meat helps the brain to function. Bradley Peterson, director of the Institute for the Developing Mind at Children's Hospital Los Angeles in California, has investigated why low iron levels in children are correlated with lower IQ and poor concentration<sup>8</sup>. Using magnetic resonance imaging, Peterson and his colleagues mapped out what happened in the brains of newborn infants of 40 adolescent mothers — a group known to be at high risk for iron deficiency. Although most of the women reported taking prenatal vitamins with iron, 58% had iron levels below normal and 14% met the criteria for mild anaemia.

As the brain develops, says Peterson, neurons become increasingly complex, forming branch-like dendrites covered with spines — much like a growing tree. The brain images that his team took showed a correlation between neuron complexity in an infant and the amount of iron in the mother's diet. "The higher the iron intake throughout pregnancy, the more mature or the more complex grey matter was at the time of birth," says Peterson, who is continuing to track the mothers and babies to see how those variations play out.

Beyond simple measures of micronutrient intake, individual requirements are also influenced by a person's genetics. So far, much of the research has focused on how people process omega-3 fatty acids, chiefly DHA and eicosapentaenoic acid (EPA), which are crucial for human cognitive health.

Omega-3 fatty acids are found primarily in oily, wild fish, such as salmon and tuna, but pasture-raised animals are also a good source. (Animals fed only soya or maize have fewer omega-3s.) In 2012, researchers discovered that

most African populations, but not European populations, carried a variant of the *FADS* gene that made them more efficient at converting omega-3s in plants into a usable form, meaning that they required less from animal sources<sup>9</sup>. Conversely, a 2014 paper reported that people carrying a variant of the *APOE* gene (11–17% of US individuals of European descent) that confers a greater risk of developing late-onset Alzheimer's disease, derived little benefit from eating fatty fish<sup>10</sup>. "One size does not fit all around nutritional recommendations," says Hosking. Put another way, the nutrients found in meat are important for health and cognition, but only up to a point. "Meat packs a lot of minerals and vitamins in just a small amount of food," says Domínguez-Rodrigo. "Eating meat is like eating a power bar."

So the key question becomes how much meat should a cognitive-health-conscious person eat. Too little can delay development and cognition. But too much, particularly if it is low quality and mass produced, is associated with other health concerns, such as heart disease and cancer, along with memory problems later in life. A person's life stage matters: pregnant women need more iron, as do babies and children. Genetics also play a part, but we don't yet know all the particulars. All these caveats make for a murky takeaway. ■

**Sujata Gupta** is a freelance science writer based in Burlington, Vermont.

- Domínguez-Rodrigo, M., Rayne Pickering, T., Semaw, S. & Rogers, M. J. *J. Hum. Evol.* **48**, 109–121 (2005).
- Neumann, C., Bwibo, N. O. & Sigman, M. *Final Report Phase II: Functional Implications of Malnutrition, Kenya Project. Nutrition CRSP.* (University of California, Los Angeles, 1992).
- Neumann, C. G., Murphy, S. P., Gewa, C., Grillenberger, M. & Bwibo, N. O. *J. Nutr.* **137**, 1119–1123 (2007).
- Hulett, J. L. *et al. Br. J. Nutr.* **111**, 875–886 (2014).
- Hosking, D. E., Nettelbeck, T., Wilson, C. & Danthiir, V. *Br. J. Nutr.* **112**, 228–237 (2014).
- Hosking, D. & Danthiir, V. *Br. J. Nutr.* **110**, 2069–2083 (2013).
- Domínguez-Rodrigo, M. *et al. PLoS ONE* **7**, e46414 (2012).
- Monk, C. *et al. Pediatric Res.* <http://dx.doi.org/10.1038/pr.2015.248> (2015).
- Mathias, R. A. *et al. PLoS ONE* **7**, e44926 (2012).
- Chouinard-Watkins, R. & Plourde, M. *Nutrients* **6**, 4452–4471 (2014).





Activities that build relationships are good for the brain.

#### SOCIAL NETWORKS

# Better together

*Social ties go hand-in-hand with cognitive health. Now researchers are trying to determine why engaging with others helps to keep the brain healthy.*

BY CHELSEA WALD

When Laura Fratiglioni looked at the data her team had collected on ageing and cognitive decline, she noticed something odd: after the age of 80, women in Stockholm were more likely than men to get dementia. Other research revealed the same trend outside of Sweden, too. Could it be that women's brains were more prone to dementia than men's, or was there some other aspect of their post-80 lifestyles causing the decline?

When they delved into the data, Fratiglioni and her colleagues found an irony: that women's well-known longevity was also their undoing. "Women are more frequently alone when they reach the age of 80 or 85 because women are more often married to older men — and in general, men die earlier than women," says Fratiglioni, who is a neurologist and director of the Aging Research Center of the Karolinska Institute in Stockholm. And, as Fratiglioni's team concluded in its seminal paper<sup>1</sup>, it is this isolation that is correlated with increased risk of dementia.

Since this work in the 1990s, the link has grown clearer. Several longitudinal studies, as well as brain studies, have shown that social ties are associated with better cognition. It is as though our friends and family can make us brighter, perhaps by stimulating thinking and reducing stress. If that is true, then working on our social networks when we are young could pay off later, by delaying both the normal decline that accompanies ageing and the pathological deterioration associated with dementia. Researchers are trying to work out which aspects of our networks do the most good, and what mechanisms link our social lives and our biology.

#### RICH RELATIONSHIPS

Over the past two decades, scientists have found associations between richer social ties — multiple links with friends, family and the community — and a variety of positive health outcomes, from reduced susceptibility to the common cold to greater life expectancy.

Initially, Fratiglioni and colleagues were uncertain whether their findings were real or whether they could be explained by reverse

causation. That is, because participants were followed for only three years on average, the team may have missed subtle, early signs of dementia that were already affecting relationships at the outset, meaning that when dementia became apparent it would seem to have been caused by degraded networks. But, in a later analysis<sup>2</sup> volunteers were followed for longer. Those who developed the disease did so about six years after monitoring began, making it less likely that undetected symptoms were biasing the results. The authors reached a similar conclusion: that the social component of any leisure activity protected against dementia as much as the mental and physical components did.

Other groups have also found support for a causal link. A six-year study in the United States by a team at the Harvard School of Public Health in Boston, Massachusetts, found that social integration — assessed by marital status, volunteer activity and frequency of contact with family and neighbours — helped to delay memory loss in older people<sup>3</sup>. The team found that memory among the least integrated declined twice as fast as it did for the most integrated. And they found no evidence of reverse causation.

The link could also apply to younger people. An analysis of survey data collected from 35- to 85-year-old Americans found that, at any age, people with more contacts and social support performed better on tests of executive function and memory<sup>4</sup>. Although this conclusion is debated, the researchers suggest that any relationship is likely to be reciprocal — cognitive function affects social engagement and vice versa.

#### DIGGING DEEP

Because studies use a variety of measures for both social network and cognitive function, the reason for the link has been hard to pin down. Almost by definition, people with strong social networks tend to have more access to information, resources, and assistance and advice from other people than do those who are isolated. But social support — even if it is well meaning — can also have negative effects on well-being, for example when it is perceived to be intrusive or controlling. "The degree and quality with which we have our social interactions impacts our entire brain and body," says cognitive psychologist Timothy Verstynen at Carnegie Mellon University in Pittsburgh, Pennsylvania.

Perhaps that shouldn't be surprising. Social networks have existed long before sites such as Facebook, and it seems that they have been an essential part of being human since our brains began tripling in size some 2 million years ago. The social-brain hypothesis, proposed by anthropologist Robin Dunbar of the University of Oxford, UK, states that primates' disproportionately large brains evolved to handle the complex demands of social living, with human brains being the most disproportionate of all.



The link between social network, cognition and brain size occurs not only at the species level, but also at the individual level. Dunbar, working with psychologist James Stiller at Nottingham Trent University, UK, found a correlation between the size of a person's network and their performance on tests of both memory and 'theory of mind' — the ability to understand another person's thoughts<sup>5</sup>. Dunbar, with US and UK colleagues, also found that the grey-matter volume of parts of the prefrontal cortex vary with social-network size, as well as with performance on theory-of-mind tasks<sup>6</sup>. The prefrontal cortex is essential for social cognition: it handles information processing, planning, working memory, language and attention. Further research by Dunbar and others revealed a standardization to human social groups, from an intimate support clique of 3–5 individuals to a broad active network of 150 people (see 'Social animals').

A similar link has been found for other brain regions. The volume of the amygdala, the almond-shaped emotion centre deep in the brain, correlates with the size and complexity of a person's social network. And grey-matter density in certain parts of the temporal lobe, which is associated with social perception and associative memory, has been found to vary according to the size of volunteers' Facebook networks<sup>7</sup>. Some researchers, including Fratiglioni, suspect that the cognitively demanding act of socializing can actually build up the brain — like exercising builds up muscles. This 'brain reserve' may then act as a buffer against functional loss, even in the face of conditions such as Alzheimer's disease.

Whether this social stimulation can form the basis of a medical intervention is under investigation. Last year, researchers led by gerontologist and biostatistician Hiroko Dodge of Oregon Health & Science University in Portland hooked up internet-based face-to-face communications systems in volunteers' homes. Using the system, participants, whose average age was just over 80, spoke with trained interviewers for 30 minutes a day for 6 weeks. After the intervention, volunteers performed better in language-based tests of executive function, for example, when they were asked to name as many words as possible that belonged to a certain category<sup>8</sup>. "Surprisingly, we found a big effect," Dodge says.

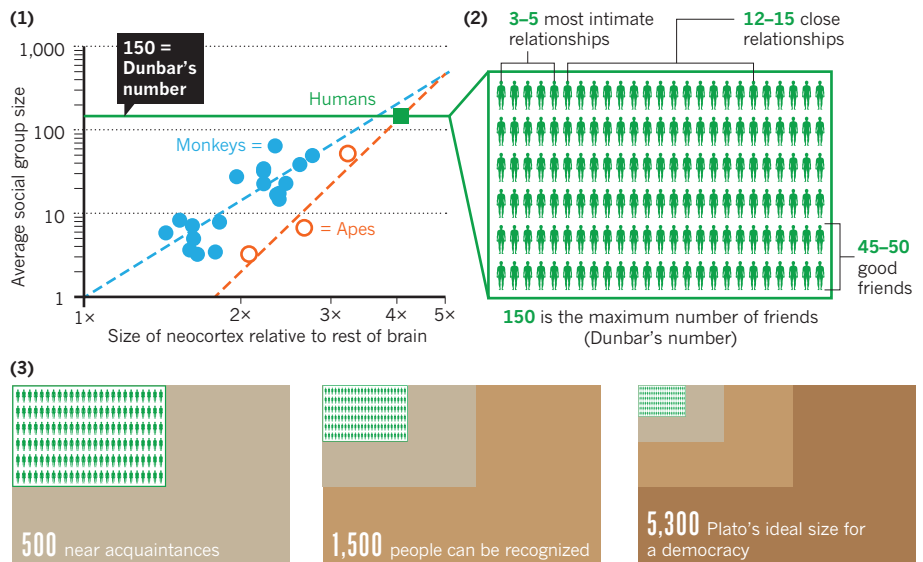
Dodge hopes to run a larger and longer study, with the goal of showing that such short conversations might delay the onset of dementia. She is optimistic about the uptake — should this simple system prove successful. Unlike physical and cognitive exercises that require special effort, she says, "talking with people is so natural".

### INFLAMED THINKING

There is, however, more to the brain than the number of little grey cells. White matter functions like "wires that connect brain areas together", supporting many aspects of cognition, says Verstynen. White-matter integrity is

## SOCIAL ANIMALS

In primates, the size of social circles correlates to the relative size of the neocortex compared with the rest of the brain (1). In humans, Dunbar's number is a suggestion of the amount of relationships a person can maintain. Human social networks seem to be split into hierarchies (2, 3), with group sizes varying by a factor of about three.



known to correlate with cognitive performance, and, in 2014, Verstynen's team revealed a link between white-matter integrity and the richness of a person's web of social interactions<sup>9</sup>. "Social-network diversity is affecting the efficiency of those wires," says Verstynen.

To explain the link between social networks and white matter, Verstynen turns to inflammation. Isolated individuals have higher levels of inflammation than those who live in a social milieu; this inflammation is comparable with that of those who smoke or who are obese. "The greater your inflammation in the body, the more you inflame myelin in the brain," Verstynen says, referring to the sheaths that protect nerves, "and sometimes that leads to degradation of the myelin."

If this is true, Verstynen says, then isolation could create a feedback loop: a reduction in social network diversity could raise levels of inflammation, damaging the white matter. That could lead to poor decision making, which in turn could lead to further shrinkage of the social network. In other words, losing friends could alter brain biology in a way that leads to losing even more friends. This might sound bleak, but "there is hope," says Verstynen. He suggests that interventions that increase someone's social network diversity or reduce systemic inflammation "might be able to break this circuit and improve brain health".

Online networks may help to decrease isolation, particularly for those with limited mobility. In a study of older people living alone,

medical sociologist Shelia Cotten of Michigan State University found that using the Internet helped to reduce feelings of loneliness and increase connections with friends and family<sup>10</sup>. Social ties, including online ones, have the potential to "provide emotional support and also provide informational support that help people make important life decisions," she says.

Sites such as Facebook could be as important for a healthy lifestyle as the gym, so long as it helps to build and maintain a strong network. "I tell people, it's very important, especially when we are getting older, to continue to be active in all three components — that is, mental, physical and social," Fratiglioni says. Five years ago, she decided to follow her own advice and took up tango. The dance style is "unbelievable," she says — "very social, very emotional, very physical". In other words, it's the perfect medicine. ■

**Chelsea Wald** is a freelance science writer based in Vienna.

1. Fratiglioni, L., Wang, H. X., Ericsson, K., Maytan, M. & Winblad, B. *Lancet* **355**, 1315–1319 (2000).
2. Wang, H.-X., Karp, A., Winblad, B. & Fratiglioni, L. *Am. J. Epidemiol.* **155**, 1081–1087 (2002).
3. Ertel, K. A., Glymour, M. M. & Berkman, L. F. *Am. J. Public Health* **98**, 1215–1220 (2008).
4. Seeman, T. E. et al. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **66**, 1141–1152 (2011).
5. Stiller, J. & Dunbar, R. I. M. *Soc. Netw.* **29**, 93–104 (2007).
6. Lewis, P. A., Rezaie, R., Brown, R., Roberts, N. & Dunbar, R. I. M. *Neuroimage* **57**, 1624–1629 (2011).
7. Kanai, R., Bahrami, B., Roylance, R. & Rees, G. *Proc. R. Soc. Lond.* **279**, 1327–1334 (2012).
8. Dodge, H. H. et al. *Alzheimers Dement.* **1**, 112 (2015).
9. Molesworth, T., Sheu, L. K., Cohen, S., Gianaros, P. J. & Verstynen, T. D. *Soc. Cogn. Affect. Neurosci.* **10**, 1169–1176 (2015).
10. Cotten, S. R., Anderson, W. A. & McCullough, B. M. *J. Med. Internet Res.* **15**, e39 (2013).



## NEURAL MODELLING

# Abstractions of the mind

*Before data were so abundant, computer models of the brain were simple. Information is now much more plentiful — but some argue that models should remain uncomplicated.*

BY KELLY RAE CHI

The first major results of the Blue Brain Project, a detailed simulation of a bit of rat neocortex about the size of a grain of coarse sand, were published last year<sup>1</sup>. The model represents 31,000 brain cells and 37 million synapses. It runs on a supercomputer and is based on data collected over 20 years. Furthermore, it behaves just like a speck of brain tissue. But therein, say critics, lies the problem. “It’s the best biophysical model we have of any brain, but that’s not enough,” says Christof Koch, a neuroscientist at the Allen Institute for Brain Science in Seattle, Washington, which has embarked on its own large-scale brain-modelling effort. The trouble with the model is that it holds no surprises: no higher functions or unexpected features have emerged from it.

Some neuroscientists, including Koch, say that this is because the model was not built with a particular hypothesis about cognitive processes in mind. Its success will depend on whether specific questions can be asked of it. The irony, says neuroscientist Alexandre Pouget, is that deriving answers will require drastic simplification of the model, “unless we figure out how to adjust the billions of parameters of the simulations, which would seem to be a challenging problem to say the least”. By contrast, Pouget’s group at the University of Geneva, Switzerland, is generating and

testing hypotheses on how the brain deals with uncertainty in functions such as attention and decision-making.

There is a widespread preference for hypothesis-driven approaches in the brain-modelling community. Some models might be very small and detailed, for example, focusing on a single synapse. Others might explore the electrical spiking of whole neurons, the communication patterns between brain areas, or even attempt to recapitulate the whole brain. But ultimately a model needs to answer questions about brain function if we are to advance our understanding of cognition.

## FROM TOP TO BOTTOM

Blue Brain is not the only sophisticated model to have hit the headlines in recent years. In late 2012, theoretical neuroscientist Chris Eliasmith at the University of Waterloo in Canada unveiled Spaun, a whole-brain model that contains 2.5 million neurons (a fraction of the human brain’s estimated 86 billion). Spaun has a digital eye and a robotic arm, and can reason through eight complex tasks such as memorizing and reciting lists, all of which involve multiple areas of the brain<sup>2</sup>. Nevertheless, Henry Markram, a neurobiologist at the Swiss Federal Institute of Technology in Lausanne who is leading the Blue Brain Project, noted<sup>3</sup> at the time: “It is not a brain model.”

Although Markram’s dismissal of Spaun

amused Eliasmith, it did not surprise him. Markram is well known for taking a different approach to modelling, as he did in the Blue Brain Project. His strategy is to build in every possible detail to derive a perfect imitation of the biological processes in the brain with the hope that higher functions will emerge — a ‘bottom-up’ approach. Researchers such as Eliasmith and Pouget take a ‘top-down’ strategy, creating simpler models based on our knowledge of behaviour. These skate over certain details, instead focusing on testing hypotheses about brain function.

Rather than dismiss the criticism, Eliasmith took Markram’s comment on board and added bottom-up detail to Spaun. He selected a handful of frontal cortex neurons, which were relatively simple to begin with, and swapped them for much more complicated neurons — ones that account for multiple ion channels and changes in electrical activity over time. Although these complicated neurons were more biologically realistic, Eliasmith found that they brought no improvement to Spaun’s performance on the original eight tasks. “A good model doesn’t introduce complexity for complexity’s sake,” he says.

## SIMPLIFY, SIMPLIFY, SIMPLIFY

For many years, computational models of the brain were what theorists call unconstrained: there were not enough experimental data to

MARIO WAGNER





map onto the models or to fully test them. For instance, scientists could record electrical activity, but from only one neuron at a time, which limited their ability to represent neural networks. Back then, brain models were simple out of necessity.

In the past decade, an array of technologies has provided more information. Imaging technology has revealed previously hidden parts of the brain. Researchers can control genes to isolate particular functions. And emerging statistical methods have helped to describe complex phenomena in simpler terms. These techniques are feeding newer generations of models.

Nevertheless, most theorists think that a good model includes only the details needed to help answer a specific question. Indeed, one of the most challenging aspects of model building is working out which details are important to include and which are acceptable to ignore. “The simpler the model is, the easier it is to analyse and understand, manipulate and test,” says cognitive and computational neuroscientist Anil Seth of the University of Sussex in Chichester, UK.

An oft-cited success in theoretical neuroscience is the Reichardt detector — a simple, top-down model for how the brain senses motion — proposed by German physicist Werner Reichardt in the 1950s. “The big advantage of the Reichardt model for motion detection was that it was an algorithm to begin with,” says neurobiologist Alexander Borst of the Max Planck Institute of Neurobiology in Martinsried, Germany. “It doesn’t speak about neurons at all.”

When Borst joined the Max Planck Society in the mid-1980s, he ran computational simulations of the Reichardt model, and got surprising

results. He found, for instance, that neurons oscillated when first presented with a pattern that was moving at constant velocity — a result that he took to Werner Reichardt, who was also taken aback. “He didn’t expect his model to show that,” says Borst. They confirmed the results in real neurons, and continued to refine and expand Reichardt’s model to gain insight into how the visual system detects motion.

In the realm of bottom-up models, the greatest success has come from a set of equations developed in 1952 to explain how flow of ions in and out of a nerve cell produces an axon potential. These Hodgkin–Huxley equations are “beautiful and inspirational,” says neurobiologist Anthony Zador of Cold Spring Harbor Laboratory in New York, adding that they have allowed many scientists to make predictions about how neuronal excitability works. The equations, or their variants, form some of the basic building blocks of many of today’s larger brain models of cognition.

### GAMBLE IN DETAILS

Although many theoretical neuroscientists do not see value in pure bottom-up approaches such as that taken by the Blue Brain Project, they do not dismiss bottom-up models entirely. These types of data-driven brain simulations have the benefit of reminding model-builders what they do not know, which can inspire new experiments. And top-down approaches can often benefit from the addition of more detail, says theoretical neuroscientist Peter Dayan of the Gatsby Computational Neuroscience Unit at University College London. “The best kind of modelling is going top-down and bottom-up simultaneously,” he says.

Borst, for example, is now approaching the Reichardt detector from the bottom up to explore questions such as how neurotransmitter receptors on motion-sensitive neurons interact. And Eliasmith’s more complex Spaun has allowed him to do other types of experiment that he couldn’t before — in particular, he can now mimic the effect of sodium-channel blockers on the brain.

Also taking a multiscale approach is neuroscientist Xiao-Jing Wang of New York University Shanghai in China, whose group described a large-scale model of the interaction of circuits across different regions of the macaque brain<sup>4</sup>. The model is built, in part, from his previous, smaller models of local neuronal circuits that show how neurons in a group fire in time. To scale up to the entire brain, Wang had to include the strength of the feedback between areas. Only now has he got the right data — thanks to the burgeoning field of connectomics (the study of connection maps within an organism’s nervous system) — to build in this important detail, he says. Wang is using his model to study decision-making, the integration of sensory information and other cognitive processes.

In physics, the marriage between experiment and theory led to the development of unifying principles. And although neuroscientists might hope for a similar revelation in their field, the brain (and biology in general) is inherently more noisy than a physical system, says computational neuroscientist Gustavo Deco of the Pompeu Fabra University in Barcelona, Spain, who is an investigator on the Human Brain Project. Deco points out that equations describing the behaviour of neurons and synapses are non-linear, and neurons are connected in a variety of ways, interacting in both a feedforward and a feedback manner. That said, there are examples of theory allowing neuroscientists to extract general principles, such as how the brain balances excitation and inhibition, and how neurons fire in synchrony, Wang says.

Complex neuroscience often requires huge computational resources. But it is not a want of supercomputers that limits good, theory-driven models. “It is a lack of knowledge about experimental facts. We need more facts and maybe more ideas,” Borst says. Those who crave vast amounts of computer power misunderstand the real challenge facing scientists who are trying to unravel the mysteries of the brain, Borst contends. “I still don’t see the need for simulating one million neurons simultaneously in order to understand what the brain is doing,” he says, referring to the large-scale simulation linked with the Human Brain Project. “I’m sure we can reduce that to a handful of neurons and get some ideas.”

Computational neuroscientist Andreas Herz, of the Ludwig-Maximilians University in Munich, Germany, agrees. “We make best progress if we focus on specific elements of neural computation,” he says. For example, a single cortical neuron receives input from thousands of other cells, but it is unclear how it processes this information. “Without this knowledge, attempts to simulate the whole brain in a seemingly biologically realistic manner are doomed to fail,” he adds.

At the same time, supercomputers do allow researchers to build details into their models and see how they compare to the originals, as with Spaun. Eliasmith has used Spaun and its variations to see what happens when he kills neurons or tweaks other features to investigate ageing, motor control or stroke damage in the brain. For him, adding complexity to a model has to serve a purpose. “We need to build bigger and bigger models in every direction, more neurons and more detail,” he says. “So that we can break them.” ■

**Kelly Rae Chi** is a freelance science writer based in Cary, North Carolina.

1. Markram, H. *et al.* *Cell* **163**, 456–492 (2015).
2. Eliasmith, C. *et al.* *Science* **338**, 1202–1205 (2012).
3. Sanders, L. *Science News* <http://go.nature.com/j1t8lu> (2012).
4. Chaudhuri, R., Knoblauch, K., Gariel, M.-A., Kennedy, H. & Wang, X.-J. *Neuron* **88**, 419–431 (2015).





## NEUROBIOLOGY

# Rise of resilience

*Stress can have a negative influence on the human brain, but increasingly it is the ability to withstand severe stress that is the focus of research.*

BY ANTHONY KING

Daniela Kaufer has a personal interest in the effects of stress. “My mum’s family had a very traumatic experience when their mother died in childbirth,” she explains. The three children grew up motherless, in 1950s war-torn Israel, but there was a marked difference in how well the siblings coped. “My mum had an extremely difficult early life,” she says. “Yet she is extremely resilient.” Kaufer, who is a neuroscientist at the University of California, Berkeley, says that why her mother in particular coped so well has fascinated her.

Research into how people react to early trauma began in earnest after the Second World War. Distressing events such as the death of a parent have been found to increase children’s short-term risk of major depression, anxiety disorders and post-traumatic stress disorder (PTSD). With advances in techniques to study genes and to explore the brain, the neurobiological study of stress is undergoing a revolution — and our view of the stress response is changing. Until about 20 years ago, the absence of a severe negative reaction such as PTSD was thought to be a lack of response. Instead, “resilience is now viewed as a reactive response”, says Kaufer.

What resilience means in terms of gene expression, numbers of cells and brain networks is now the focus of research. “Years ago, most would have thought that resilient individuals escape some of the bad things that

stress induces in the brains of more susceptible individuals,” explains neurobiologist Eric Nestler at Mount Sinai School of Medicine, New York. “Now we believe susceptible individuals lack some of the more adaptive changes that occur in the resilient brain.”

Stress is an unavoidable component of life, and the stress response is a crucial survival mechanism. “The brain is a detector of threatening information,” says neuropsychologist Sonia Lupien at the University of Montreal, Canada. “This is its most important job if you want to survive.” Acute stress readies us for action, but chronic stress wears us down, altering the brain genetically and neurologically and priming us for mental health problems. Whether a person is resilient to stress depends heavily on their life history. Understanding the effects of early-life difficulties could provide new ways to treat or prevent mental illnesses such as severe depression or PTSD in susceptible individuals.

### STRESS IN THE BRAIN

Confronted with a life-threatening situation, hormones and neurotransmitters prep us for action. Specific stress hormones — cortisol in primates, corticosterone in most rodents — are released, some of which surge across the blood–brain barrier. Stress gets everywhere: all our cells host receptors for the hormone. “Every brain area has something happen to it,” says Kaufer. The human brain has two types of receptor for cortisol. One has a six to tenfold higher affinity for the

molecule than the other, and so is activated earlier, by smaller amounts of cortisol.

The hippocampus (which is pivotal for memory) and the amygdala (the centre for emotions) contain lots of the high-affinity receptors, and are, therefore, activated by slight rises in the hormone. The frontal lobe, which is involved in executive planning and control, has only the low-affinity receptor, and is activated later, after the tide has risen. And, as Lupien and colleagues found, both memory formation and recall in adults can be influenced by cortisol<sup>1</sup>.

The existence of two receptor types means that response to stress is not linear. “The relationship between circulating stress hormone and memory is an inverted U-shape function,” Lupien explains. “Up to a certain level, stress hormones are good for your memory” — when the cortisol binds only to the high-affinity receptors, the ability to lay down and retrieve memory is enhanced. When the low-affinity receptors are activated, the relationship enters the right-hand side of the U-shape and the response shifts, she adds.

The duration of stress is also important. A transient bout of stress causes a proliferation of neural stem cells and a spike in numbers of new neurons, which take at least two weeks to mature. The brain seems to be preparing itself in case a second stressor comes calling. Chronic stress is not so beneficial. It slashes investment in new neurons, prunes the tree-like shape of existing ones, and suppresses new connections.

If stress hormones remain elevated for

MARIO WAGNER

months or years, they can stimulate physiological changes: the hippocampus shrinks and the amygdala grows, for example. Eventually, the complex feedback system that suppresses the excess secretion of cortisol is disturbed. Once this happens, the capacity to discriminate between threat levels falls away. Either everything seems threatening (anxiety) or else nothing does (depression or burnout).

### EARLY INFLUENCES

Old ideas that certain individuals have an inherent 'hardiness' or an innate ability to bounce back from severe stress have fallen by the wayside. Instead, resilience and our response to trauma are recognized as being more dynamic, changing throughout life. It's a complicated milieu, but one of the main ways that stress marks the brain is through epigenetics. This does not change genes, but it can change their expression by attaching methyl groups to DNA or associated proteins.

At McGill University in Montreal, neuroscientist Michael Meaney's group has been exploring the stress response in rats. It found that high-licking and low-licking grooming strategies in mother rats gave rise to different offspring<sup>2</sup>. The lesser-grooming mums produced pups with high anxiety, poor stress-recovery and low cognitive performance. The pups' brain circuits that switch off stress were sluggish, owing to higher DNA methylation and lower expression of the 'off' receptors in the hippocampus. Well-groomed pups showed the opposite.

People might be tempted to label high-groomers as better mothers, says Kaufer. "But it is not 'good mums' or 'bad mums', just a different parenting style." Parenting style can reflect the environment and prepare the offspring, she explains. Being a cautious, worried rat — the offspring of a less-thorough groomer — makes sense if you live in an alley full of cats.

"The stress response is one of the most conserved things in evolution," Kaufer says. This means that animal findings tend to be applicable to humans. However, what is good for avoiding predators may not be a healthy adaptation to the continued stress of modern life. Kieran O'Donnell, a neuroscientist in Meaney's lab, says that the epigenetic changes in the anxious rat brains seem to have human analogies. "We see the same sorts of changes in DNA methylation of the hormone receptor in people who suffered child maltreatment," he says. "The question is, can we intervene and do anything about this?" O'Donnell is currently investigating whether the children of vulnerable young mums who received regular nurse visits after they had given birth show changes in DNA methylation 27 years later.

Chronic stress in early life stamps an especially long-lasting mark on the brain. Some people carry an epigenetic signature of stress from exposure as a baby — or even as an embryo. Emotional stress in pregnant women,



A white 'bully mouse' intimidates a smaller mouse, which then develops signs of social withdrawal.

for example, can alter their children's epigenetics and the neural connections in their babies' amygdalas. Later exposure to severe trauma can activate this signature, explains neurobiologist Alon Chen, of the Weizmann Institute of Science in Rehovot, Israel. He was involved in a

*"The stress response is one of the most conserved things in evolution."*

study that reported that chronic stress can reduce the number of hormone receptors in the hippocampus needed to shut off the stress response<sup>3</sup>. This means that individuals exposed to chronic stress early in life may be more susceptible to a stress-related disorder as an adult, owing to a disrupted feedback loop. "It is not healthy to stay in a stress situation for a long time," says Chen. "Returning to a normal level is an important part of the stress-response machinery."

### BEATING THE BULLIES

Stress affects our relationships with others (and socializing is itself an agent in cognitive health; see page S14). Nestler has devised a 'bully mouse' scenario (pictured) in which, for 5 to 10 minutes a day over 10 days, a normal mouse is placed in a cage that is already occupied by a larger, more aggressive strain of mouse that intimidates the incomer. At all other times, the mice are kept close enough to see and smell one another, but they are separated by mesh. Nestler's team found that afterwards, some of the bullied mice avoid all social contact, even with smaller, non-aggressive mice.

As with memory, the way that sociability changes with stress is not linear. Kaufer's lab found that rats exposed to moderate stress — in this case, being immobilized in a bag — displayed more positive social behaviour, such as huddling, resource sharing and reduced aggression<sup>4</sup>. The researchers also saw an increase in the prosocial hormone oxytocin. But if the immobilized rats were exposed to fox odour, the addition of this high-level-stress inducer caused them to lose all pro-social behaviours. Oxytocin plummeted, as did its

receptors. "This is really interesting because it can start to explain the social withdrawal that you can see in some psychopathologies like PTSD and depression," Kaufer says.

Yet even in these scenarios, some rodents did better than others. Nestler and his colleagues found that some of their mice were resilient to the bullying, and that these mice showed a greater number of gene-expression changes<sup>5</sup>. So resilience was a reaction, endowing the individuals with greater adaptability. "We don't know the underlying cause," says Nestler, adding that the search is on for genes with altered expression in the brains of resilient individuals. His lab is now planning to look at the role of individual messenger RNAs and proteins in mediating resilience to try and tease out what is different in the cells and neural circuits of resistant mice.

This research could deliver benefits for treating stress-related disorders such as depression. Most antidepressant drug-discovery efforts have focused on ways to undo the bad effects of stress. "Understanding resilience offers an additional approach," says Nestler, "to look for ways to induce mechanisms of natural resilience in those individuals who are inherently more susceptible."

Research into stress is changing how we view mental-health conditions. Epigenetic and brain-chemistry changes caused by life stresses can be reversed by activities such as exercise, yoga, meditation and mental stimulation. And soon these types of behavioural intervention might be complemented by pharmaceuticals. "We may learn the molecular mechanisms important for resilience," says O'Donnell. "And then use that to help those susceptible to stress or who have suffered ill treatment or trauma." ■

**Anthony King** is a freelance science writer based in Dublin.

1. Lupien, S. et al. *J. Neurosci.* **74**, 2893–2903 (1994).
2. Weaver, I. C. G. et al. *Nature Neurosci.* **7**, 847–854 (2004).
3. Taliaz, D. et al. *J. Neurosci.* **31**, 4475–4483 (2011).
4. Muroy, S. E., Long, K. L., Kaufer, D. & Kirby, E. D. *Neuropsychopharmacology* <http://dx.doi.org/10.1038/npp.2016.16> (2016)
5. Krishnan, V. et al. *Cell* **131**, 391–404 (2007).



# Progress and challenges in probing the human brain

Russell A. Poldrack<sup>1</sup> & Martha J. Farah<sup>2</sup>

Perhaps one of the greatest scientific challenges is to understand the human brain. Here we review current methods in human neuroscience, highlighting the ways that they have been used to study the neural bases of the human mind. We begin with a consideration of different levels of description relevant to human neuroscience, from molecules to large-scale networks, and then review the methods that probe these levels and the ability of these methods to test hypotheses about causal mechanisms. Functional MRI is considered in particular detail, as it has been responsible for much of the recent growth of human neuroscience research. We briefly review its inferential strengths and weaknesses and present examples of new analytic approaches that allow inferences beyond simple localization of psychological processes. Finally, we review the prospects for real-world applications and new scientific challenges for human neuroscience.

The way that we conceptualize brain function has always been constrained by the methods available to study it. Studies of patients with focal brain lesions in the nineteenth century led to the view of the brain as a collection of focal centres specialized for particular cognitive abilities, such as ‘Broca’s area’ for speech production. The development of neurophysiological recording techniques in the twentieth century led to Barlow’s ‘neuron doctrine’, according to which the functions of individual neurons can be extrapolated to explain the function of the brain as a whole. The cognitive neuroimaging studies of the 1980s focused on subtractive comparisons between cognitive tasks meant to isolate specific cognitive operations, and led to a relatively modular view of brain function as involving localized and separable regions that implement elementary mental operations.

The methods of contemporary human neuroscience have provided a much more complex and nuanced view of the human brain as a dynamic network with multiple levels of organization, in which function is characterized by a balance of regional specialization and network integration. Although current methods are limited in their utility for studying brain function at fine-grained levels of organization (such as single neurons or cortical columns), human neuroscience has nonetheless made remarkable progress in understanding basic aspects of functional organization, and with this have come a number of applications to address real-world problems. Our goal here is to review the current state of human neuroscience, focusing on what kinds of questions can and cannot be answered using current techniques and how those answers are relevant to real-world applications.

## How can we study the human brain?

Methods for studying human brain function can be organized according to the kinds of mechanistic insights that each technique provides. As shown in Table 1 the first characteristic is the level of mechanism captured by the method. Mechanisms range from the molecular level (neurotransmitters and receptors) to large-scale networks (the dynamic integration and coordination of different functional areas of the brain). Although this distinction is related to physical scale, it does not depend on the method’s spatial resolution per se. For example, positron emission tomography (PET) using neurotransmitter ligands measures molecular mechanisms, even though its spatial resolution is on the order

of one centimetre. The second characteristic is the ability of each method to elucidate the mechanistic role of an observed brain molecule, cell, region or network in a mental function of interest. By mechanism we mean the causal chain of events that result in the realization of a function. To fully understand human brain function is to know the causal chains of events at the molecular, cellular, population, and network levels that give rise to psychological function. For this reason, the power to identify causal relationships is a crucial dimension of difference among methods.

Some methods used in the study of human brain function provide relatively little insight into causal mechanisms. This includes methods that exploit naturally occurring variation by observing the strength of association between individual differences in brain function and behaviour. Analysis of relationships between behavioural traits, genes, brain structure, and brain function exemplify this approach (see Box 1 for a discussion of genomic approaches). For many important psychological phenomena, from effects of life history to personality traits, we are limited to observational methods. For example, individual differences in the personality trait of impulsiveness have been associated with differences in striatal dopamine release<sup>1</sup>, functional MRI (fMRI) activation<sup>2</sup>, and cortical grey matter volume<sup>3</sup>. Observed associations between neural and psychological traits do not necessarily imply a causal relationship, as these associations could result from an unmeasured third variable that independently influences the two measures. Nevertheless, such associations provide a valuable starting point for theorizing about the neural mechanisms of human psychology, and their evidentiary value can be strengthened by measuring possible confounds to rule them in or out.

Although functional neuroimaging, electroencephalography/magnetoencephalography (EEG/MEG) and single-cell recordings are sometimes criticized as being purely correlative and therefore uninformative about mechanism, that criticism is only partly accurate. When psychological processes are experimentally manipulated by presenting a certain kind of stimulus and/or engaging the subject in a task, we can infer that any reliably elicited brain activity was caused by performing these psychological functions. We cannot, however, infer with confidence that the observed brain activity is causally responsible for the psychological process under study. Despite this limitation (which is shared by neuronal recordings in non-human animals), neuroimaging studies

<sup>1</sup>Department of Psychology, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Center for Neuroscience & Society, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.



**Table 1 | An overview of the levels of analysis and levels of causal inference afforded by different human neuroscience methods**

		Level of mechanism			
		Molecules	Cells	Populations	Networks
Strength of causal evidence	Purely observational (associations do not necessarily imply causal relations between mind and brain)	Genetic associations with behaviour, brain function or brain structure Postmortem studies of gene expression Correlations of MRI spectroscopy or PET ligand imaging with psychological traits		Structural morphometry correlated with psychological traits	Resting functional connectivity (fMRI, EEG/MEG) or structural connectivity (sMRI, DTI) correlated with psychological traits
	Manipulate psychological process and observe brain (neural measures may be epiphenomenal)	Task modulation studies using PET with neurotransmitter ligands or MRI spectroscopy	Intracerebral recording in surgical patients	Task activation studies (PET, fMRI, EEG/MEG) Representational analysis (fMRI, EEG/MEG) Computational neuroimaging (fMRI, EEG/MEG)	Task-based functional connectivity (fMRI, EEG/MEG)
	Manipulate brain and observe psychological results (demonstrates causal effect of neural system in behaviour)	Pharmacological manipulation (including hormones)	Direct brain stimulation in surgical patients	Focal cortical lesions Transcranial magnetic stimulation Transcranial electrical stimulation Cortical surface electrode stimulation in surgical patients	Disconnection/white matter lesions

DTI, diffusion tensor imaging; EEG/MEG, electroencephalography/magnetoencephalography; fMRI, functional MRI; MRI, magnetic resonance imaging; PET, positron emission tomography; sMRI, structural MRI.

in which psychological processes are manipulated comprise the majority of current human neuroscience research, and have advanced our understanding of human brain function, as we will discuss in more detail below.

### BOX 1

## Challenges of merging neuroimaging and genomics

The substantial heritability of many psychological functions has driven great interest in finding genetic underpinnings of individual differences in neural function. Twin and family studies have demonstrated significant heritability for both task-related BOLD responses<sup>91</sup> and resting-state functional connectivity<sup>92</sup> in fMRI. In the past decade, a large number of studies have also reported associations between BOLD signals and common variants in candidate genes. Unfortunately, this approach has generally been unsuccessful in identifying genetic associations that are replicated in genome-wide association studies (GWAS). For example, a striking finding from the first well-powered GWAS of genetic variants associated with brain volume was that none of the associations previously identified through candidate gene studies were replicated at the genome-wide level<sup>88</sup>. Similarly, candidate gene associations with cognitive function (such as the association between polymorphisms in the *COMT* gene and working memory) and brain activation have generally not been confirmed in meta-analyses, and are subject to a substantial degree of publication bias<sup>93,94</sup>. Like for many other areas of genetics, this suggests that genome-wide approaches are the most likely to lead to reliable identification of common variants related to brain structure and function. However, GWAS approaches require large samples (in the tens of thousands) which are very difficult to amass for task-based fMRI studies; for that reason, GWAS-based approaches to probing the human brain will likely be limited to resting-state fMRI and structural MRI. Other strategies, such as targeted studies investigating rare variants of large effect identified using genome sequencing or studies using gene expression in peripheral tissues may have greater utility for genetic studies of task-based fMRI. Task-based fMRI may also be used to further investigate candidate variants identified on the basis of GWAS studies of psychiatric disorders or population variability.

More decisive evidence concerning causal necessity can be obtained by manipulating the brain itself to assess the resulting effect on the psychological process in question. Naturally occurring or surgical lesions, which provided the basis for most of what we knew about human brain function before the advent of neuroimaging, are still of great interest because they provide insight into the causal necessity of specific brain regions or connections. More recently developed methods of brain stimulation allow for reversible inhibition or excitation of a brain area, thereby expanding our ability to test the causal role of brain regions in the mechanisms of human thought and action. Deep brain stimulation (DBS) provides the most precise method for targeted stimulation by using surgically implanted electrodes, but is limited to situations where patients are undergoing implantation for medical reasons. Use of non-invasive brain stimulation for research purposes has grown rapidly in recent decades, starting with transcranial magnetic stimulation (TMS), in which pulsed magnetic fields induce currents in the brain. Various forms of transcranial electric stimulation (TES), in which current is delivered using external electrodes, have also been used, of which the most common variant is transcranial direct current stimulation (tDCS). Unlike DBS, non-invasive brain stimulation generally affects larger and more superficial areas of the brain, but researchers are seeking to improve spatial resolution with new magnetic coil shapes for TMS and new electrode configurations for tDCS. Focused ultrasound is also being explored as a means to stimulate more precisely delimited brain regions<sup>4</sup>. Pharmacological agonists and antagonists of particular neurotransmitter systems can be used to experimentally manipulate the human brain at the molecular level, although with imperfect specificity<sup>5</sup>. By combining each of these manipulations of brain function with functional brain imaging, one can leverage the causal information obtained through pharmacological challenges or brain stimulation. For example, the causal role of activity in specific brain regions, identified using fMRI, for a particular function has been tested by brain stimulation, using both direct cortical stimulation (for example, ref. 6) and TMS<sup>7</sup>.

### New capabilities of fMRI

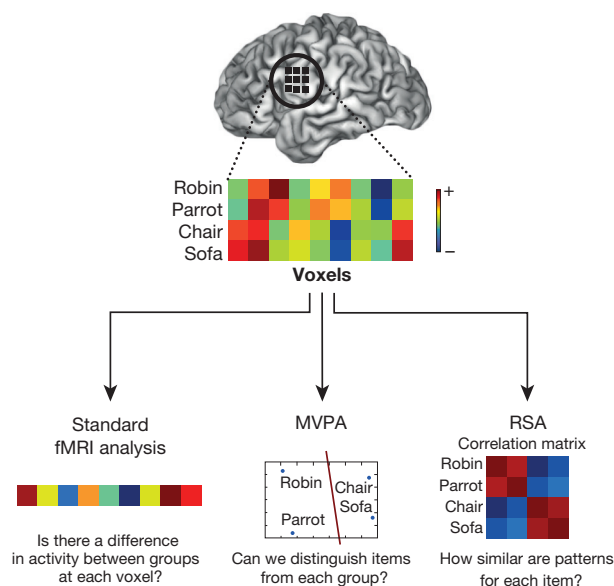
Because fMRI has become the main method for the study of human brain function, our review focuses on this method and new ways of using it. In the last two decades, fMRI has transitioned from a newly developed technique for revealing neuronal activity to being the workhorse method of cognitive neuroscience (see the recent special issue of *Neuroimage* on

the first 20 years of fMRI<sup>8</sup>). Much has been learned about the biological mechanisms underlying blood oxygen level dependent (BOLD) signals<sup>9,10</sup>, but still much remains to be understood, such as the roles of specific glial and neuronal cell types in the coupling of neuronal activity to blood flow (for example, refs 11, 12). This limited physiological understanding poses problems for the interpretation of fMRI data. In particular, although fMRI signals often correlate strongly with both action potentials ('spikes') and local field potentials, they are largely reflective of post-synaptic processes, and in some cases they can be dissociated from spiking altogether<sup>13</sup>. The relative sensitivity of fMRI to post-synaptic processes as opposed to spiking has been seen as a drawback by some who view spikes as the essence of brain function, but it is worth noting that this discovery has actually rekindled interest in the analysis of local field potentials in electrophysiology (where these signals have long been discarded) (for example, ref. 14), and suggests that fMRI may sometimes be sensitive to subthreshold signals that would be missed by analysis of spikes only. Uncertainties in relating fMRI to psychological, as well as physiological, processes have also been debated, and progress has been made on this front too. From experimental approaches such as adaptation paradigms for probing representations to analyses of functional connectivity, fMRI is routinely used to answer questions about mind–brain relationships that go far beyond localization<sup>15</sup>. Here we discuss three examples of new approaches to understanding human brain function with fMRI that address questions of representation, computational processes and network interactions across the brain.

## Representational analyses

Early work in neuroimaging focused largely on 'brain mapping'—identifying regions based on the mental processes that cause them to be activated. This approach has provided a large body of reliable associations between function and structure, but has not been particularly successful in providing new insights into how psychological functions are implemented<sup>16</sup>. However, two relatively recent approaches, known as multi-voxel pattern analysis (MVPA)<sup>17</sup> and representational similarity analysis (RSA)<sup>18</sup>, can more directly relate psychological contents to brain function (Fig. 1). MVPA involves the use of methods from the field of machine learning to decode or predict psychological states from patterns of brain activation across voxels (hence the term 'brain-reading'). Since its introduction more than a decade ago, MVPA has been used in a number of domains to demonstrate the predictive ability of fMRI activation patterns. Perhaps the most impressive are demonstrations of the ability to successfully reconstruct visual scenes<sup>19</sup> and faces<sup>20</sup> from BOLD activity patterns; similar advances have been made for higher cognitive functions such as word meaning<sup>21</sup>. These studies go beyond simply differentiating between experimental conditions, as they show how the underlying representational spaces relate to brain activity; for example, using a related approach known as voxel-wise modelling, Huth and colleagues<sup>22</sup> developed a model that estimated the response at each location on the cortical surface to a large number of visual and semantic features present in natural movies (Fig. 2). MVPA approaches have also provided new insights into the neural organization of cognitive functions. For example, MVPA has informed our understanding of the mechanisms of visual attention, by showing that attention changes both the representation of stimuli across regions of visual cortex as well as the mutual information between regions<sup>23</sup>. In the domain of memory, MVPA has been used to show that competition between memory representations in working memory leads to poorer subsequent memory for those items, demonstrating a nonmonotonic relationship between competition and subsequent memory<sup>24</sup>.

Whereas MVPA is generally used to decode individual psychological states, RSA instead asks how the patterns of brain activity evoked by different stimuli are related to one another, and thus provides the means to directly address questions of how mental representations are implemented in the brain. RSA has enabled the demonstration of

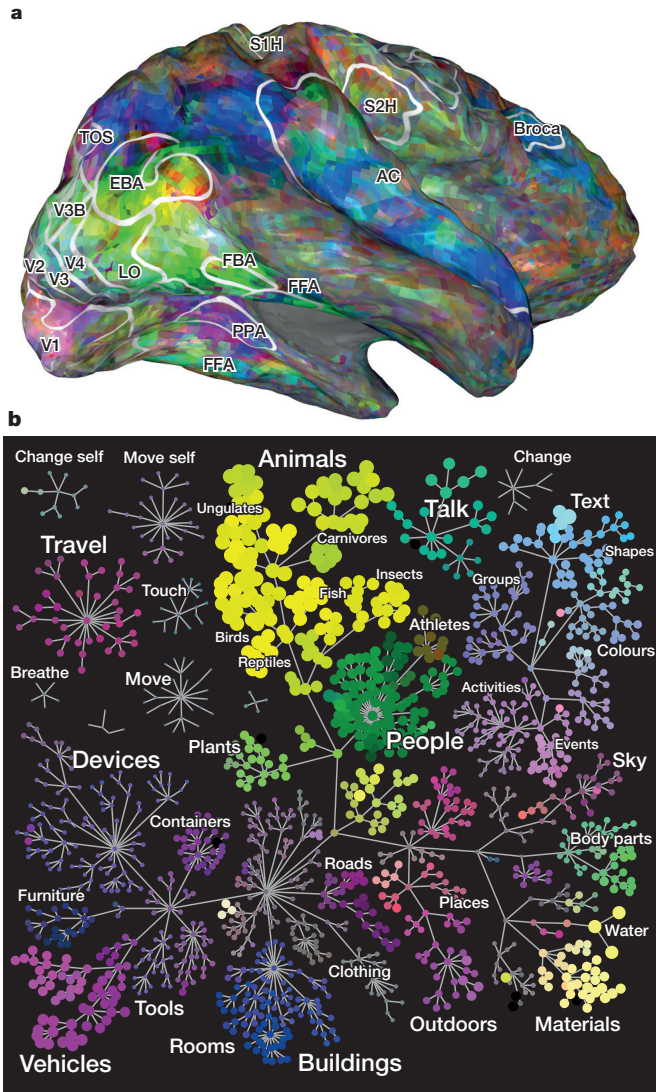


**Figure 1 | Different approaches to the analysis of fMRI data.** This example depicts data from a hypothetical study in which four different stimuli were presented (two birds and two items of furniture) and response measured for each item across nine voxels; intensity of activity is depicted from blue (negative) to red (positive). The standard univariate fMRI analysis approach would examine the difference at each voxel between the averages of the two categories. Multi-voxel pattern analysis (MVPA) examines the multidimensional relationship between patterns of activity, in this case projecting the nine-dimensional space of voxel patterns (the voxel vector) into a two-dimensional space and identifying a boundary that separates items from the two classes. Representational similarity analysis (RSA) examines the correlations between activity patterns for each item, in this case showing that items within category show a high correlation (red), whereas the correlation of items between categories is low (blue).

direct isomorphisms between psychological representations of stimuli (such as the similarity or typicality of objects) and the neural patterns associated with those stimuli<sup>25,26</sup>. Because psychological theories often make predictions regarding the similarity of different stimuli, RSA has also enabled the direct testing of theories, such as theories about how categories are represented<sup>27</sup> and theories of how repeated experiences lead to enhanced learning<sup>28</sup>. RSA can be applied to any kind of multi-dimensional data, and this has enabled the demonstration of systematic mappings of visual object representations between humans (using fMRI) and non-human primates (using electrophysiological recordings)<sup>29</sup>—an example that highlights how human neuroscience can also help to establish more direct parallels with findings in non-human models, allowing insights to filter in both directions.

Although much MVPA and RSA work (as depicted in Fig. 1) has focused on the representations found in localized brain regions, these methods are equally useful for assessing representations that are spread across the brain. For example, recent work has shown that mental states such as physical pain can be decoded by analysis of patterns of activation across brain regions<sup>30</sup>.

The legitimate enthusiasm about these methods is tempered by lingering questions regarding the interpretation of multivariate analyses<sup>31,32</sup>. In addition, recent work combining electrophysiology and fMRI in non-human primates has demonstrated that the sensitivity of MVPA is limited by the spatial characteristics of the neuronal representations that code for particular features, such that some kinds of neuronal patterns may be more difficult to decode using MVPA than others<sup>33</sup>. Finally, it is important to stress that, like standard neuroimaging approaches, MVPA and RSA approaches do not inform about causal mechanisms.



**Figure 2 | A mapping of high-dimensional semantic space onto the cortical surface.** Here, voxel patterns for 1,705 different action and object categories, based on brain activity obtained during viewing of natural movies<sup>22</sup> are mapped onto the cortical surface image generated using online browser at (<http://gallantlab.org/semanticmovies/>). **a**, Mapping of semantic categories to each point on the surface; the colours on the surface map correspond to the semantic map in panel **b**. **b**, A depiction of the semantic space derived from all semantically selective voxels. Categories that have similar colours in the semantic space are represented in similar patterns of voxels in the brain. Data from ref. 22.

### Integrating fMRI and computational modelling

Computational models play a central role in our understanding of both cognitive and brain functions and, increasingly, of the relationship between the two. By making assumptions explicit, computational models enable more direct testing of theories, as well as providing the means to link computations at the neuronal level with higher-order functions. An example of an area in which substantial progress has been made using this approach is reinforcement learning, in which an animal selects actions and learns from the rewards gained from those actions. Computational models of reinforcement learning (RL) have long played a central role in artificial intelligence and psychology, and the discovery by Schultz and colleagues<sup>34</sup> that dopamine neurons appear to signal one of the important quantities in these models (reward prediction error) has brought these models to the forefront of the neuroscience of decision making. For example, a set of publications in 2003 applied RL models to neuroimaging data and thereby identified correlates of reward predic-

tion error signals in dopaminergic target regions such as the ventral striatum<sup>35,36</sup>. Subsequent neuroimaging work has established that there are multiple RL signals in the brain, some reflecting the simple association between actions and values (known as 'model-free' RL) and others reflecting more complex contextual and hierarchical learning processes (known as 'model-based' RL)<sup>37,38</sup>. Similarly, in the study of memory, progress has been made in the mapping of medial temporal lobe subregions to specific computational operations such as pattern completion and pattern separation (for example, ref. 39). In each of these domains, the computational interpretation of neuroimaging signals has been greatly enhanced by parallel studies in non-human animals, allowing imaging signals to be linked more directly to direct measures of neuronal activity.

### Functional connectivity analysis and resting-state fMRI

Perhaps the most revolutionary development to arise from human neuroimaging research is the realization that the resting brain is far from quiescent, and that important insights into brain function can be gained from studying the correlated fluctuations of signals across the brain at rest. Much of the research into the resting state has focused on a set of regions (including anterior and posterior midline regions, lateral temporoparietal cortex, and the medial temporal lobe, known as the 'default mode' network<sup>40</sup>) that are consistently less active during performance of difficult tasks<sup>41</sup>, and are functionally connected in the resting state<sup>42</sup>. Similar patterns of resting connectivity have been observed in non-human primates<sup>43</sup> and awake rodents<sup>44</sup>, suggesting that they reflect fundamental principles of mammalian brain organization. There is also growing evidence that these networks may be important in brain disorders. For example, the posterior portion of the default mode network appears to play a critical role in the memory deficits observed in Alzheimer disease, showing a convergence of amyloid deposition, structural atrophy, and decreased metabolic activity<sup>45</sup>.

Data collected in the resting state can provide insights into the broader functional organization of the brain as well. In particular, the organization of resting state signals bears a close relation to the organization of brain activity evoked by mental tasks. For example, Smith *et al.*<sup>46</sup> used independent component analysis to identify spatially independent sets of voxels from resting-state fMRI data and from task-based data (obtained from the Brainmap meta-analytic database), and demonstrated that the components extracted from resting-state fMRI showed a high degree of concordance with those extracted from task-based data. The overlap between resting-state and task-based functional organization can also be seen within individuals; for example, the longitudinal examination of a single individual revealed reliable spatial parcellation of activity in the cerebral cortex (using resting fMRI data) that mapped systematically to the activation patterns observed across a large number of task measurements<sup>47</sup>.

Despite the substantial excitement around resting-state fMRI findings, numerous concerns have been raised about their interpretation. In particular, there are lingering questions regarding the ways in which artefacts related to head motion and physiological fluctuations may influence estimates of resting state connectivity, and whether common data analytic methods may induce systematic artefacts<sup>48,49</sup>. In addition, potential confounds such as light sleep<sup>50</sup> may drive differences in resting state signals. The unconstrained nature of resting-state fMRI is a double-edged sword; it is potentially very useful for the study of clinical groups for whom task performance may be difficult, but at the same time, it is not possible to determine whether group differences reflect fundamental differences in functional connectivity or relative differences in the ongoing mental content of different groups during rest (see ref. 51).

### Applications of human neuroscience

With the development of new methods have come attempts to apply them to real-world problems, in both medical and non-medical contexts. (See Box 2 for a discussion of the ethical, legal, and societal issues raised by these applications.)



## BOX 2

## Ethical, legal and societal impact of human neuroscience

As the methods of human neuroscience find broader application, they affect human life in new ways. The field of 'neuroethics' is concerned with ethical, legal and societal issues raised by these new applications<sup>95</sup>.

Two kinds of problems have emerged from the increasing ability of brain imaging to reveal aspects of individual psychology: problems that arise from the current and imminent capabilities of these methods, and problems that arise from their lack of claimed capabilities. To the extent that imaging can predict important personal characteristics such as health status, academic achievement, and criminal behaviour, its use must be managed with care to protect privacy and avoid discrimination<sup>96</sup>. To the extent that imaging cannot provide help with high-stake problems, the public should be protected from claims that it can. For example, a seemingly 'scientific' method for detecting lies or diagnosing psychiatric disorders<sup>66,97</sup> has a strong appeal to the general public who cannot be expected to appreciate the gap between what is claimed and what is established fact.

New ways of changing brain function pharmaceutically, and with electromagnetic stimulation, also raise new ethical issues. Of course, humanity has long manipulated brain function to modify mental states using substances such as alcohol and caffeine. However, psychopharmacology has broadly penetrated our everyday lives and the scope of psychiatric diagnoses and treatment has expanded—a societal shift that some find troubling<sup>98</sup>. Furthermore, many now use psychoactive drugs purely for enhancement of healthy brain function rather than to treat a medical condition<sup>99</sup>. Aside from issues of safety and efficacy, brain enhancement raises issues of fairness (is it akin to doping in sports?), justice (will the ability to access enhancements widen the already existing gaps between haves and have-nots?) and social standards (will unenhanced job performance become sub-standard?).

Non-invasive brain stimulation is the newest method for brain enhancement. Simple transcranial electrical stimulation (for example, tDCS) devices are available to consumers at relatively low cost and regulation is minimal<sup>100</sup>. Given the public interest in this method and the rudimentary state of knowledge about its effects, it is crucial that the safety and efficacy of these methods are established. The efficacy of cognitive enhancement with tDCS is hotly debated<sup>101</sup> and whether long-term use of tDCS is safe has yet to be studied. In addition, neuroethical issues of fairness, justice and social standards mentioned above also apply to enhancement of brain function by brain stimulation.

## Brain disorders

The methods of human neuroscience hold particular promise for understanding and treating psychiatric disorders, because these disorders do not have clear analogues in non-human animals, and animal models currently used for preclinical screening of potential therapies are increasingly regarded as being inadequate<sup>52</sup>. In the absence of valid animal models, it becomes all the more crucial to apply new methods for understanding human brain function and dysfunction. The goal of improving the treatment of neuropsychiatric disorders is made even more challenging because of our current diagnostic system. Although depression, schizophrenia, autism and other serious psychiatric disorders have long been considered disorders of the brain, they are still diagnosed exclusively by behavioural signs and symptoms. These diagnostic criteria do not seem to have clear relations to the biological processes that would be targeted by new medical treatments.

In response to this problem, an alternative way of systematizing psychiatric disorders has been developed—the NIMH Research Domain

Criteria (RDoC)<sup>53</sup>—that describes disorders according to impairments in specific functional systems of the brain (such as fear or reward learning) and at different levels of mechanism of the kind represented in Table 1 (for example, molecules or circuits). RDoC characterizations cut across traditional diagnostic categories and are intended to capture the underlying pathophysiology more accurately. Given the multiple levels of mechanism captured by the RDoC, the system encourages research with a broad array of methods to identify potentially targetable dysfunctions.

The application of several human neuroscience methods has led to the development of targeted treatments, for example, in the field of depression. Functional imaging studies have highlighted the role of the subgenual anterior cingulate cortex in a network of regions involved in mood, leading Mayberg and colleagues to use deep brain stimulation in this area to regulate mood in depressed patients<sup>54</sup>. Lateral prefrontal regions, implicated through imaging studies in depression, have been targeted with non-invasive brain stimulation, including the FDA-approved use of TMS for treatment-resistant depression. Functional neuroimaging can itself be used as a treatment, by providing patients with a real-time measure of regional brain activity to use as a biofeedback signal. This approach is being tested for the treatment of chronic pain, depression and addiction<sup>55</sup>. In contrast, neuroimaging has not so far been very successful in aiding differential diagnosis of disorders in terms of current diagnostic categories. A recent large meta-analysis identified a set of regions in which structural abnormalities were consistently associated with psychiatric disorders, but found very little specificity for individual disorders<sup>56</sup>, consistent with the notion that current diagnostic distinctions are not biologically realistic categories.

Another approach to the discovery of therapeutic targets is the use of genetic association studies to identify sets of genes that are associated with a disorder and that together may indicate particular molecular pathways underlying the disorder. Although the numbers of subjects needed to establish reliable genetic associations is daunting, progress has been made through large international collaborations. For example, Psychiatric Genomics Consortium has to date identified more than 100 common genetic variants reliably implicated in schizophrenia<sup>57</sup>. Imaging can also be used to develop endophenotypes (or intermediate phenotypes) that may bear a closer relation to the effect of a gene variant than does disease diagnosis, as well as to mitigate the problem of heterogeneity within conventional diagnostic categories (see Box 1).

It may be less surprising that the methods developed for human neuroscience research have been applied in the diagnosis and treatment of neurological diseases, but at least two recent developments deserve mention here. Studies of Alzheimer disease at mechanistic levels from molecules to systems have improved diagnostic accuracy and have enabled a degree of prediction before clinical signs of the disease<sup>58</sup>. Molecular biomarkers from blood and CSF, and patterns of brain activity and structure have revolutionized clinical research in this area by facilitating trials of preventive treatment and by providing intermediate phenotypes as early gauges of effectiveness. Disorders of consciousness following severe brain damage are another area of clinical neuroscience for which neuroimaging shows promise. Some patients who have been diagnosed as being in the vegetative state can follow commands to imagine actions that activate specific areas of the brain in much the same way as healthy control subjects do, and can even use these imagined actions to answer questions (for example, "Do you have any brothers? If yes, imagine playing tennis, if no, imagine walking through your house.")<sup>59</sup>. Thus, neuroimaging offers new insights into the assessment of consciousness, as well as the distinct problem of prognosis, in severely brain-damaged patients.

## Predicting behaviour

The ability to predict future behaviour is of value in almost every sphere of human activity. Although it has often been said that "the best predictor of future behaviour is past behaviour," in some cases brain imaging can improve our ability to predict future behaviour, over and

above what we can do with behavioural history. Marketing professionals were among the first to attempt to predict behaviour using brain imaging. Recognizing the limitations of focus groups and other traditional methods to discern what consumers want, they have used functional neuroimaging to predict the effects of different advertising campaigns, packaging, and other factors on consumer behaviour, based on the premise that activity in the brain's reward or motivation centres may be a more direct measurement of wanting than are verbal self-reports<sup>60</sup>. Although most of this work is conducted by and for corporations aiming to improve sales rather than share scientific knowledge, published academic studies have begun to lend some credence to the potential of neuromarketing. For example, when teenage subjects were scanned while listening to unfamiliar songs, the reward system activity evoked by the songs, but not the subjects' ratings of their likeability, was predictive of sales of the songs over the subsequent three years<sup>61</sup>.

Prediction is also important outside of business. Falk and colleagues have adapted neuromarketing methods for the purpose of creating more effective public service announcements. They showed that brain responses (but not ratings) to an anti-smoking advertisement were predictive of subsequent call volume to an anti-smoking hotline<sup>62</sup>. Gabrieli *et al.*<sup>63</sup> recently summarized evidence concerning neuroimaging-based prediction in domains ranging from healthful eating to criminal recidivism, including numerous examples of prediction of educational outcomes. Indeed, neuroimaging can predict future academic skills over and above traditional behavioural predictors, thus enabling earlier and more appropriate interventions to address individual children's reading and math difficulties. These authors also pointed out a number of methodological challenges in neuroimaging-based prediction of behaviour, including the need to develop and test predictions with different samples, to avoid the 'optimism bias' that occurs when predictions are tested in the same population from which they were generated.

### Human neuroscience in the courtroom

In recent years the methods of human neuroscience have found their way into the courtroom. Perhaps the most obvious, but also the most misunderstood, role for neuroscience is in helping to determine criminal responsibility. Proving that a criminal act may have had a neural cause is not in itself exculpatory, as every human act is caused by the brain<sup>64</sup>. However, to the extent that neuroscience can provide evidence of mental dysfunction (for example, a tumour in the frontal cortex that may have impaired the ability to control behaviour), immaturity or other psychological grounds for reduced criminal responsibility, it is potentially relevant and has been used. For example, the Supreme Court explicitly cited neuroscience evidence in its decision in *Graham v. Florida* to abolish life in prison without parole for juveniles who commit non-homicidal offences. It is more difficult to make legal arguments for applying neuroimaging evidence to individual cases because most findings from neuroimaging research are generalizations based on groups of people and may therefore not allow reliable inferences regarding individuals<sup>65</sup>. Nevertheless, neuroimaging scans from defendants are sometimes presented in the sentencing phase of criminal trials as grounds for mitigation of the sentence, as weaker evidentiary standards apply in the sentencing phase.

Neuroimaging can be applied in ways other than determining degree of responsibility. Lie detection is one example that has been pursued in legal contexts, although it has not so far been admitted into US courts and has yet to demonstrate validity, reliability or resistance to countermeasures outside of the laboratory<sup>66</sup>. Another application concerns pain: brain-based biomarkers for pain would help discriminate real suffering from malingering—a pivotal issue in many lawsuits—and have been admitted as evidence in at least one US case<sup>67</sup>.

### Challenges and future directions for neuroimaging

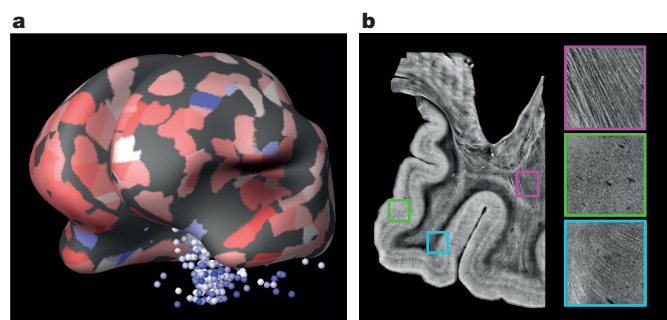
The field of neuroimaging is growing rapidly, and there are a number of exciting new directions on the horizon.

### New technologies for imaging and manipulating the human brain

Rapid advances in non-human neuroscience have been driven by the development of technologies that measure and manipulate brain function with increasing precision. Human neuroscience has lagged in this respect, in part because of the ethical challenges associated with direct manipulation and neuronal recording of the human brain. However, in response to the urgent need for better treatments for psychiatric disorders, research is underway with the aim to design implantable systems for sensing and modulating human brain networks<sup>68</sup>. The development of optogenetic and 'opto-fMRI' approaches in non-human primates<sup>69</sup> suggests that these methods may one day become feasible for use in human studies, and it is likely that electrical brain stimulation will eventually be supplemented with optogenetic approaches. Although such invasive techniques will likely only be used in rare clinical cases (that is, patients are undergoing implantation for medical reasons), they have the potential to provide much greater specificity in circuit mapping.

fMRI will probably remain the principal neuroimaging method in humans in the foreseeable future. However, the ongoing BRAIN initiative in the United States<sup>70</sup> is providing substantial funding to develop entirely new techniques for imaging of brain function, and a significant proportion of this funding will go specifically towards the development of new methods for imaging the human brain. In addition, new developments in MRI have greatly increased the utility of standard MRI systems. For example, multiband imaging techniques<sup>71</sup> have enabled a several-fold increase in the temporal resolution of fMRI acquisitions, and higher MRI field strengths (7 tesla and higher) hold promise to enable improvements in spatial resolution as well (for example, ref. 72). There is thus great reason to be optimistic that methodological limits will continue to be pushed in the future.

Additional insight into human brain function will likely come from the study of postmortem human brains, which has long been a staple method for the characterization of anatomical structure and study of brain disorders. New techniques have enhanced the ability to visualize the structure of human brain tissue (Fig. 3). For example, optical coherence tomography has been used to image *ex vivo* human cortical tissue, providing high-resolution imaging of cytoarchitecture with less distortion than standard microscopy techniques<sup>73</sup>. The first whole-brain atlas of genome-wide gene expression in postmortem human brains<sup>74</sup> has provided an important resource for understanding how gene expression relates to brain function; for example, the maps from this project have



**Figure 3 | New methods for characterizing the postmortem human brain.**

**a**, A map of expression of the serotonin receptor 3B displayed on the reconstructed cortical surface in one individual from the Allen Brain Atlas Human Brain data set (generated using data from <http://human.brain-map.org/>). **b**, Optical coherence tomography imaging of the human brain (2.9 in-plane resolution). Large panel presents an average intensity projection in depth over 300; inset zooms are maximum intensity projections over 300, showing fibres in the white matter (pink inset), fibres arcing through the subcortical junction to insert into the cortex (cyan inset), and neurons in the cortex (bright spots in the green inset). Image courtesy of Bruce Fischl, Caroline Magnain and David Boas, Massachusetts General Hospital.

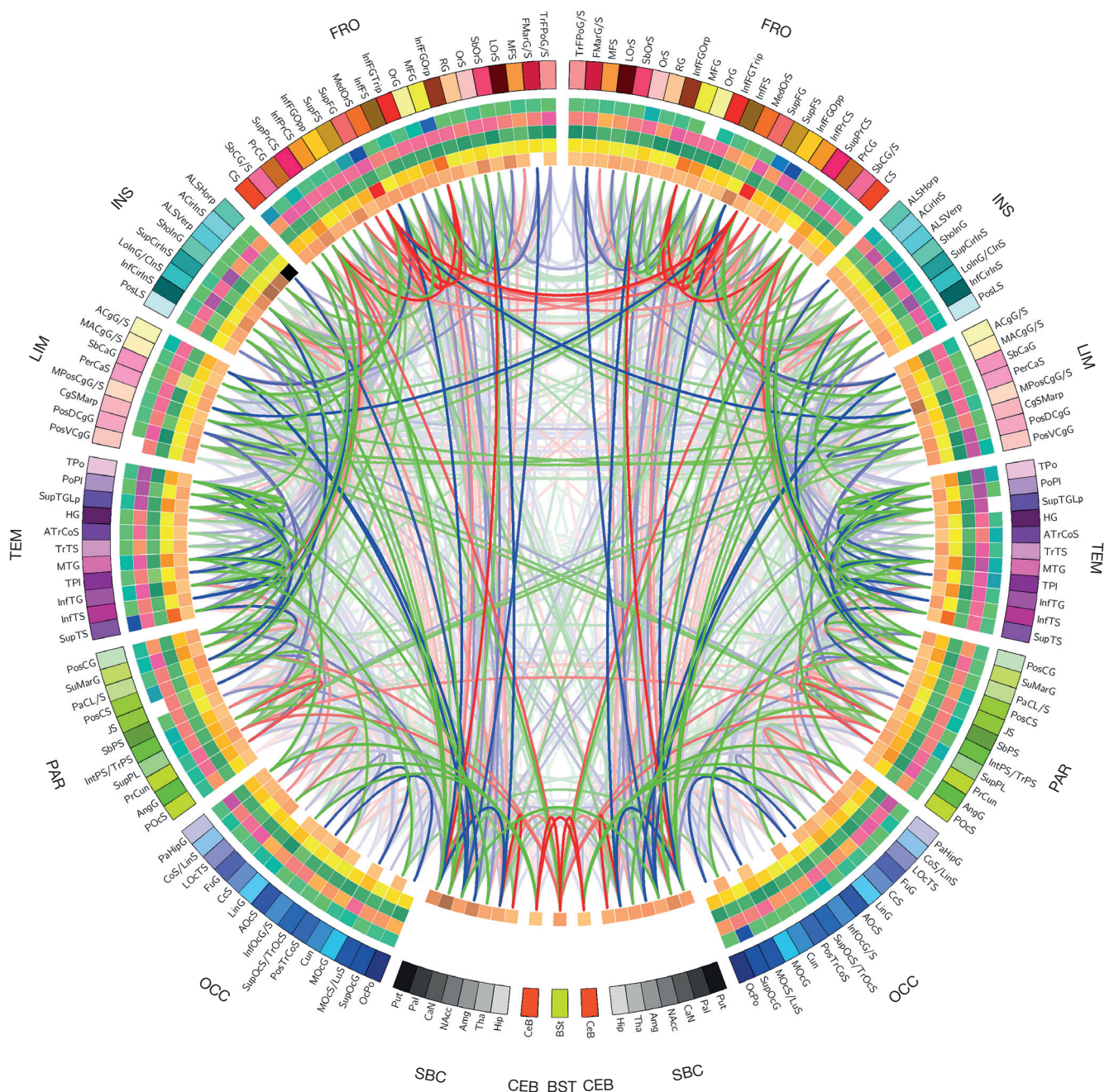


been used to identify expression differences across different resting-state networks<sup>75</sup>. Continued development of such resources will be essential for progress in understanding the genetic architecture of brain function and their relation to mental health disorders.

### Connectomics

The Human Connectome Project<sup>76</sup> is nearing completion, and has already provided a rich database for the modelling of functional and anatomical connectivity of the human brain. However, fundamental challenges remain. For example, diffusion MRI provides the means to track white matter pathways (Fig. 4) and has been used to identify white

matter connectivity disruptions associated with cognitive disorders such as dyslexia<sup>77</sup>; however, diffusion imaging has inherent biases that limit its ability to accurately track connections across the entire brain<sup>78,79</sup>. The last decade has seen a proliferation of approaches to model functional connectivity on the basis of functional MRI data, though the dust has yet to settle regarding which methods are most effective (for example, ref. 80). To determine this, the analysis methods must be validated, which is challenging to do in humans but may be achieved using direct measurements of functional connectivity from invasive human approaches and non-human animals to validate the neuroimaging results. There is increasing evidence that at least in non-human primates



**Figure 4 | A 'connectogram'<sup>90</sup> for an example healthy adult female subject.** The outermost ring shows the various brain regions arranged by lobe (fr, frontal; ins, insula; lim, limbic; tem, temporal; par, parietal; occ, occipital; nc, non-cortical; bs, brain stem; CeB, cerebellum) and further ordered anterior (top) to posterior (bottom). The colour map of each region is lobe-specific and maps to the colour of each regional parcellation as determined using FreeSurfer.

The set of five rings (from the outside inward) reflect grey matter volume, area, thickness, curvature, and connectivity density. The lines inside of the circle represent the computed degrees of connectivity between segmented brain regions using diffusion tractography, with colour representing the relative fractional anisotropy of the connection (from blue to red). Image courtesy of Jack Van Horn, University of Southern California.



functional connectivity reflects anatomical connectivity as measured using either diffusion MRI<sup>81</sup> or anatomical tract-tracing<sup>82</sup>; but it remains an important challenge to establish the ways in which functional and diffusion connectivity measures converge or diverge.

### Reproducibility of neuroimaging research

Large-scale meta-analyses have made it clear that neuroimaging results can be highly convergent across studies, to the degree that cognitive processes can be accurately inferred from individual subject data using decoders trained on meta-analytic data based on reported activation coordinates<sup>83</sup>. However, the last few years have also seen increasing concern regarding the reproducibility of research findings in neuroscience, paralleling more general concerns about reproducibility of scientific results<sup>84</sup>. These issues are particularly acute for neuroimaging given the high dimensionality of the data, relatively low statistical power of many studies<sup>85</sup>, high degree of analytic flexibility in data analysis procedures<sup>86</sup>, and potential for questionable research practices such as circular analysis procedures<sup>87</sup>. The field of neuroimaging has been at the forefront of a number of developments that aim to improve reproducibility and the sharing of data are increasingly being embraced. The Alzheimer's Disease Neuroimaging Initiative (ADNI), International Neuroimaging Data Sharing Initiative (INDI), ENIGMA, and the Human Connectome Project together have shared thousands of neuroimaging data sets and this has enabled a number of novel discoveries. For example, data sharing by the ENIGMA consortium has enabled the first well-powered genome-wide association study of brain volume<sup>88</sup>, identifying replicated associations between brain volume and several common genetic variants. In addition, nearly all of the main software packages for neuroimaging data analysis are free and open source, providing transparency and reproducibility in data analysis across groups, and the publication of fully reproducible analysis workflows has begun (for example, ref. 89). The increasing use of machine learning methods, with their focus on out-of-sample generalization rather than statistical significance, is also leading to a greater emphasis on achieving reproducibility.

### Outlook

The use of new tools for imaging and manipulating the brain will continue to advance our understanding of how the human brain gives rise to thought and action. The combination of myriad methods with different and complementary strengths and weaknesses will allow neuroscientists to develop a multilevel understanding of the brain, spanning from molecules to large-scale networks. New analysis methods have advanced fMRI beyond 'blobology' and will provide direct insight into the mapping of mental and neural representations, while newer analysis and acquisition methods will offer other novel insights into the relation of mind and brain. fMRI and other human neuroscience methods will continue being applied to solve real-world problems, within medicine and beyond. Although some of these applications are currently premature relative to the demonstrated capabilities of the methods, it is clear that the new methods of human neuroscience will have much to offer science and society.

Received 31 May; accepted 4 September 2015.

1. Buckholz, J. W. *et al.* Dopaminergic network differences in human impulsivity. *Science* **329**, 532 (2010).
2. Plichta, M. M. & Scheres, A. Ventral-striatal responsiveness during reward anticipation in ADHD and its relation to trait impulsivity in the healthy population: a meta-analytic review of the fMRI literature. *Neurosci. Biobehav. Rev.* **38**, 125–134 (2014).
3. Schilling, C. *et al.* Common structural correlates of trait impulsiveness and perceptual reasoning in adolescence. *Hum. Brain Mapp.* **34**, 374–383 (2013).
4. Legon, W. *et al.* Transcranial focused ultrasound modulates the activity of primary somatosensory cortex in humans. *Nature Neurosci.* **17**, 322–329 (2014).
5. Chamberlain, S. R., Müller, U., Robbins, T. W. & Sahakian, B. J. Neuropharmacological modulation of cognition. *Curr. Opin. Neurol.* **19**, 607–612 (2006).
6. Parvizi, J. *et al.* Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* **32**, 14915–14920 (2012).

**This study uses the combination of fMRI and intracranial electrical stimulation to demonstrate the causal role of fusiform regions in face perception.**

7. Chen, A. C. *et al.* Causal interactions between fronto-parietal central executive and default-mode networks in humans. *Proc. Natl Acad. Sci. USA* **110**, 19944–19949 (2013).
  8. Bandettini, P. A. Twenty years of functional MRI: the science and the stories. *Neuroimage* **62**, 575–588 (2012).
  9. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).
  10. Attwell, D. *et al.* Glial and neuronal control of brain blood flow. *Nature* **468**, 232–243 (2010).
  11. Hillman, E. M. C. Coupling mechanism and significance of the bold signal: a status report. *Annu. Rev. Neurosci.* **37**, 161–181 (2014).
  12. Sirotni, Y. B. & Das, A. Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature* **457**, 475–479 (2009).
  13. Thomsen, K., Offenhauser, N. & Lauritzen, M. Principal neuron spiking: neither necessary nor sufficient for cerebral blood flow in rat cerebellum. *J. Physiol. (Lond.)* **560**, 181–189 (2004).
  14. Einevoll, G. T., Kayser, C., Logothetis, N. K. & Panzeri, S. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature Rev. Neurosci.* **14**, 770–785 (2013).
  15. Farah, M. J. Brain images, babies, and bathwater: critiquing critiques of functional neuroimaging. *Hastings Cent. Rep.* **44**, S19–S30 (2014).
  16. Poldrack, R. A. & Yarkoni, T. From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annu. Rev. Psychol.* <http://dx.doi.org/10.1146/annurev-psych-122414-033729> (2015).
  17. Norman, K. A., Polyn, S. M., Detre, G. J. & Haxby, J. V. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* **10**, 424–430 (2006).
  18. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers Syst. Neurosci.* **2**, 4 (2008).
  19. Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M. & Gallant, J. L. Bayesian reconstruction of natural images from human brain activity. *Neuron* **63**, 902–915 (2009).
  20. Cowen, A. S., Chun, M. M. & Kuhl, B. A. Neural portraits of perception: reconstructing face images from evoked brain activity. *Neuroimage* **94**, 12–22 (2014).
  21. Mitchell, T. M. *et al.* Predicting human brain activity associated with the meanings of nouns. *Science* **320**, 1191–1195 (2008).
- An outstanding example of the use of fMRI along with a model of word meaning derived from a large text corpus to predict activation patterns associated with words.**
22. Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76**, 1210–1224 (2012).
  23. Sprague, T. C., Saproo, S. & Serences, J. T. Visual attention mitigates information loss in small- and large-scale neural codes. *Trends Cogn. Sci.* **19**, 215–226 (2015).
  24. Lewis-Peacock, J. A. & Norman, K. A. Competition between items in working memory leads to forgetting. *Nature Commun.* **5**, 5768 (2014).
  25. Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D. & Kriegeskorte, N. Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc. Natl Acad. Sci. USA* **111**, 14565–14570 (2014).
  26. Davis, T. & Poldrack, R. A. Quantifying the internal structure of categories using a neural typicality measure. *Cereb. Cortex* **24**, 1720–1737 (2014).
  27. Mack, M. L., Preston, A. R. & Love, B. C. Decoding the brain's algorithm for categorization from its neural implementation. *Curr. Biol.* **23**, 2023–2027 (2013).
  28. Xue, G. *et al.* Greater neural pattern similarity across repetitions is associated with better memory. *Science* **330**, 97–101 (2010).
  29. Kriegeskorte, N. *et al.* Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126–1141 (2008).
- This paper applies representational similarity analysis to human fMRI and monkey electrophysiology data to demonstrate similar representational spaces in the inferior temporal cortex across species.**
30. Wager, T. D. *et al.* An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* **368**, 1388–1397 (2013).
  31. Davis, T. *et al.* What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *Neuroimage* **97**, 271–283 (2014).
  32. Todd, M. T., Nystrom, L. E. & Cohen, J. D. Confounds in multivariate pattern analysis: theory and rule representation case study. *Neuroimage* **77**, 157–165 (2013).
  33. Dubois, J., de Berker, A. O. & Tsao, D. Y. Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J. Neurosci.* **35**, 2791–2802 (2015).
  34. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
  35. McClure, S. M., Berns, G. S. & Montague, P. R. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* **38**, 339–346 (2003).
  36. O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H. & Dolan, R. J. Temporal difference models and reward-related learning in the human brain. *Neuron* **38**, 329–337 (2003).
  37. Badre, D. & Frank, M. J. Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: evidence from fMRI. *Cereb. Cortex* **22**, 527–536 (2012).

38. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
39. LaRocque, K. F. *et al.* Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J. Neurosci.* **33**, 5466–5474 (2013).
40. Raichle, M. E. The brain's default mode network. *Annu. Rev. Neurosci.* **38**, 433–447 (2015).
41. Shulman, G. L. *et al.* Common blood flow changes across visual tasks: II. decreases in cerebral cortex. *J. Cogn. Neurosci.* **9**, 648–663 (1997).
42. Greicius, M. D., Krasnow, B., Reiss, A. L. & Menon, V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl Acad. Sci. USA* **100**, 253–258 (2003).
43. Vincent, J. L. *et al.* Intrinsic functional architecture in the anaesthetized monkey brain. *Nature* **447**, 83–86 (2007).
44. Becerra, L., Pendse, G., Chang, P.-C., Bishop, J. & Borsook, D. Robust reproducible resting state networks in the awake rodent brain. *PLoS ONE* **6**, e25701 (2011).
45. Buckner, R. L. *et al.* Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *J. Neurosci.* **25**, 7709–7717 (2005).
- This paper presents a multimodal analysis implicating the default mode network in cognitive decline associated with Alzheimer disease.**
46. Smith, S. M. *et al.* Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl Acad. Sci. USA* **106**, 13040–13045 (2009).
- This paper demonstrates that resting state networks are systematically associated with cognitive functions.**
47. Laumann, T. O. *et al.* Functional system and areal organization of a highly-sampled individual human brain. *Neuron* **87**, 657–670 (2015).
48. Murphy, K., Birn, R. M. & Bandettini, P. A. Resting-state fMRI confounds and cleanup. *Neuroimage* **80**, 349–359 (2013).
49. Power, J. D., Schlaggar, B. L. & Petersen, S. E. Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage* **105**, 536–551 (2015).
50. Tagliazucchi, E. & Laufs, H. Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* **82**, 695–708 (2014).
51. Morcom, A. M. & Fletcher, P. C. Does the brain have a baseline? Why we should be resisting a rest. *Neuroimage* **37**, 1073–1082 (2007).
52. Nestler, E. J. & Hyman, S. E. Animal models of neuropsychiatric disorders. *Nature Neurosci.* **13**, 1161–1169 (2010).
53. Insel, T. R. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *Am. J. Psychiatry* **171**, 395–397 (2014).
54. Mayberg, H. S. Targeted electrode-based modulation of neural circuits for depression. *J. Clin. Invest.* **119**, 717–725 (2009).
55. Sulzer, J. *et al.* Real-time fMRI neurofeedback: progress and challenges. *Neuroimage* **76**, 386–399 (2013).
56. Goodkind, M. *et al.* Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* **72**, 305–315 (2015).
- This paper examines a large structural imaging dataset and finds that brain abnormalities linked to mental illness are shared across diagnostic categories.**
57. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
58. Sperling, R. A., Karlawish, J. & Johnson, K. A. Preclinical Alzheimer disease—the challenges ahead. *Nature Rev. Neurol.* **9**, 54–58 (2013).
59. Owen, A. M. Detecting consciousness: a unique role for neuroimaging. *Annu. Rev. Psychol.* **64**, 109–133 (2013).
60. Ariely, D. & Berns, G. S. Neuromarketing: the hope and hype of neuroimaging in business. *Nature Rev. Neurosci.* **11**, 284–292 (2010).
61. Berns, G. S. & Moore, S. A neural predictor of cultural popularity. *J. Consum. Psychol.* **22**, 154–160 (2012).
62. Falk, E. B., Berkman, E. T. & Lieberman, M. D. From neural responses to population behavior: neural focus group predicts population-level media effects. *Psychol. Sci.* **23**, 439–445 (2012).
63. Gabrieli, J. D. E., Ghosh, S. S. & Whitfield-Gabrieli, S. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* **85**, 11–26 (2015).
64. Morse, S. J. Brain overclaim syndrome and criminal responsibility: A diagnostic note. *Ohio State J. Criminal Law* **3**, 397–412 (2006).
65. Jones, O. D., Wagner, A. D., Faigman, D. L. & Raichle, M. E. Neuroscientists in court. *Nature Rev. Neurosci.* **14**, 730–736 (2013).
66. Farah, M. J., Hutchinson, J. B., Phelps, E. A. & Wagner, A. D. Functional MRI-based lie detection: scientific and societal challenges. *Nature Rev. Neurosci.* **15**, 123–131 (2014).
67. Reardon, S. Neuroscience in court: the painful truth. *Nature* **518**, 474–476 (2015).
68. Underwood, E. DARPA aims to rebuild brains. *Science* **342**, 1029–1030 (2013).
69. Gerits, A. *et al.* Optogenetically induced behavioral and functional network changes in primates. *Curr. Biol.* **22**, 1722–1726 (2012).
70. Insel, T. R., Landis, S. C. & Collins, F. S. The NIH Brain Initiative. *Science* **340**, 687–688 (2013).
71. Moeller, S. *et al.* Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn. Reson. Med.* **63**, 1144–1153 (2010).
72. Yacoub, E., Harel, N. & Ugurbil, K. High-field fMRI unveils orientation columns in humans. *Proc. Natl Acad. Sci. USA* **105**, 10607–10612 (2008).
73. Magnain, C. *et al.* Optical coherence tomography visualizes neurons in human entorhinal cortex. *Neurophotonics* **2**, 015004 (2015).
- A demonstration of the power of optical coherence tomography to image neural structure in ex vivo human brain tissue.**
74. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
75. Richiardi, J. *et al.* Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244 (2015).
76. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).
- This paper presents a broad overview of the Human Connectome Project.**
77. Saygin, Z. M. *et al.* Tracking the roots of reading ability: white matter volume and integrity correlate with phonological awareness in prereading and early-reading kindergarten children. *J. Neurosci.* **33**, 13251–13258 (2013).
78. Van Essen, D. C. *et al.* Mapping Connections in Humans and Non-human Primates: Aspirations and Challenges for Diffusion Imaging 2nd edn, Ch. 16 (Elsevier, 2013).
79. Eveley, C. *et al.* Superficial white matter fiber systems impede detection of long-range cortical connections in diffusion MR tractography. *Proc. Natl Acad. Sci. USA* **112**, E2820–E2828 (2015).
80. Smith, S. M. *et al.* Network modelling methods for fMRI. *Neuroimage* **54**, 875–891 (2011).
81. Honey, C. J. *et al.* Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl Acad. Sci. USA* **106**, 2035–2040 (2009).
82. Shen, K. *et al.* Information processing architecture of functionally defined clusters in the macaque cortex. *J. Neurosci.* **32**, 17465–17476 (2012).
83. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods* **8**, 665–670 (2011).
84. Ioannidis, J. P. A. Why most published research findings are false: author's reply to Goodman and Greenland. *PLoS Med.* **4**, e215 (2007).
85. Button, K. S. *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nature Rev. Neurosci.* **14**, 365–376 (2013).
86. Carp, J. On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments. *Frontiers Neurosci.* **6**, 149 (2012).
87. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F. & Baker, C. I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neurosci.* **12**, 535–540 (2009).
88. Stein, J. L. *et al.* Identification of common variants associated with human hippocampal and intracranial volumes. *Nature Genet.* **44**, 552–561 (2012).
- This paper presents the first well-powered genome-wide association study of brain structure.**
89. Waskom, M. L., Kumaran, D., Gordon, A. M., Rissman, J. & Wagner, A. D. Frontoparietal representations of task context support the flexible control of goal-directed cognition. *J. Neurosci.* **34**, 10743–10755 (2014).
90. Irimia, A., Chambers, M. C., Torgerson, C. M. & Van Horn, J. D. Circular representation of human cortical networks for subject and population-level connectomic visualization. *Neuroimage* **60**, 1340–1351 (2012).
91. Blokland, G. A. M. *et al.* Heritability of working memory brain activation. *J. Neurosci.* **31**, 10882–10890 (2011).
92. Glahn, D. C. *et al.* Genetic control over the resting brain. *Proc. Natl Acad. Sci. USA* **107**, 1223–1228 (2010).
93. Barnett, J. H., Scoriels, L. & Munafò, M. R. Meta-analysis of the cognitive effects of the catechol-O-methyltransferase gene Val158/108Met polymorphism. *Biol. Psychiatry* **64**, 137–144 (2008).
94. Nickl-Jockschat, T., Janouschek, H., Eickhoff, S. B. & Eickhoff, C. R. Lack of meta-analytic evidence for an impact of COMT Val158Met genotype on brain activation during working memory tasks. *Biol. Psychiatry* <http://dx.doi.org/10.1016/j.biopsych.2015.02.030> (2015).
95. Farah, M. J. Neuroethics: the ethical, legal, and societal impact of neuroscience. *Annu. Rev. Psychol.* **63**, 571–591 (2012).
96. Illes, J. & Racine, E. Imaging or imagining? A neuroethics challenge informed by genetics. *Am. J. Bioeth.* **5**, 5–18 (2005).
97. Farah, M. J. & Gillihan, S. J. The puzzle of neuroimaging and psychiatric diagnosis: technology and nosology in an evolving discipline. *AJOB Neurosci.* **3**, 31–41 (2012).
98. Conrad, P. *The Medicalization of Society: On the Transformation of Human Conditions into Treatable Disorders* (Johns Hopkins Univ. Press, 2007).
99. Sahakian, B. & Morein-Zamir, S. Professor's little helper. *Nature* **450**, 1157–1159 (2007).
100. Fitz, N. S. & Reiner, P. B. The challenge of crafting policy for do-it-yourself brain stimulation. *J. Med. Ethics* **41**, 410–412 (2015).
101. Horvath, J. C., Forte, J. D. & Carter, O. Quantitative review finds no evidence of cognitive effects in healthy populations from single-session transcranial direct current stimulation (tDCS). *Brain Stimul.* **8**, 535–550 (2015).

**Acknowledgements** Thanks to I. Eisenberg, D. Glahn, R. Raizada, and M. Shine for comments on an earlier draft of this manuscript, and N. Logothetis for helpful discussions.

**Author Contributions** R.P. and M.F. planned and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.P. ([poldrack@stanford.edu](mailto:poldrack@stanford.edu)).

## REVIEW

# Nutrition and brain aging: how can we move ahead?

P Barberger-Gateau<sup>1,2</sup>

Epidemiological studies and basic research suggest a protective effect of long-chain omega-3 polyunsaturated fatty acids, antioxidants and B vitamins against brain aging. However, most randomized controlled trial (RCTs) with nutritional supplements have yielded disappointing effects on cognition so far. This paper suggests some original directions for future research to better support a role of nutrition in brain aging. The role of other nutrients such as docosapentaenoic acid and fat-soluble vitamins D and K should be investigated. A more holistic approach of nutrition is necessary, encompassing potential synergies between nutrients as found in a balanced diet. Potential beneficiaries of a nutritional supplementation should be better targeted, according to their dietary, cognitive and maybe genetic characteristics. Innovative RCTs should be implemented to assess the impact of nutrition for the prevention or treatment of cognitive decline in older persons, using intermediate biomarkers of disease progression and mechanisms of action of nutrients as outcomes.

*European Journal of Clinical Nutrition* (2014) **68**, 1245–1249; doi:10.1038/ejcn.2014.177; published online 27 August 2014

## INTRODUCTION

In accordance with the results of basic research, epidemiological studies suggest a protective effect of several classes of nutrients against cognitive decline and risk of dementia,<sup>1,2</sup> including principally long-chain omega-3 polyunsaturated fatty acids (n-3 PUFA),<sup>3,4</sup> vitamins C and E,<sup>5</sup> carotenoids,<sup>6</sup> polyphenols<sup>7</sup> and B vitamins,<sup>8</sup> although some discordant data exist as well. Many underlying mechanisms can be evoked to support biological plausibility of a protective effect of these nutrients against brain aging: role in brain composition and function,<sup>9</sup> hippocampal neurogenesis,<sup>10</sup> limitation of accumulation of the beta-amyloid peptide,<sup>11,12</sup> decreased oxidative stress and inflammation,<sup>13</sup> decreased homocysteine concentration<sup>8</sup> and vascular effects.<sup>14</sup>

However, most well-conducted randomized controlled trials (RCTs) with nutritional supplements have yielded disappointing effects on cognition so far (for reviews, see Wald *et al.*,<sup>15</sup> Mazereeuw *et al.*<sup>16</sup> and Dangour *et al.*<sup>17</sup>). As written by Dangour *et al.*<sup>17</sup> regarding n-3 PUFA but easily transferable to other nutrients, whether discrepancies between epidemiological and intervention studies 'reflect a real absence of benefit on cognitive function from (nutritional) supplementation, or whether they reflect intrinsic limitations in the design of published studies remains open to question'. The aim of this paper was to suggest new areas of research to better support a role of nutrition in brain aging. Two aspects are developed: nutritional exposure, concerning the investigation of new nutrients and a more holistic approach of the diet, and innovative RCTs with nutritional supplements, regarding their inclusion criteria and outcomes in relation with cognition.

## INVESTIGATING NEW NUTRIENTS

There is convincing evidence that fats are a major class of nutrients for brain development, structure and functioning.<sup>9</sup>

However, previous studies have neglected some fatty acids or fat-soluble vitamins that could have an important role in the brain.

### Docosapentaenoic acid (DPA)

Actually, the term DPA refers to two different fatty acids: one in the omega3 series C22:5n-3, situated between eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) in the desaturation and elongation chain of alpha-linolenic acid, and another one in the omega6 series C22:5n-6 resulting from elongation, desaturation and beta-oxidation of arachidonic acid. As EPA and DHA, n-3 DPA is mainly found in fatty fish and could contribute to the apparently protective effect of fish consumption against cognitive decline. The metabolism and biological functions of DPA are still poorly understood. Supplementation with pure n-3 DPA induced a significant rise in the proportions of EPA and DHA in plasma triacylglycerides, suggesting that DPA may act as a reservoir of the major long-chain n-3 PUFA in humans.<sup>18</sup> In healthy humans under their spontaneous diet, n-3 DPA is found in non-negligible amounts in cerebrospinal fluid (CSF) and positively correlated with total plasma n-3 PUFA.<sup>19</sup>

N-6 DPA in erythrocytes was higher in 48 older individuals with mild cognitive impairment (MCI) than in 27 healthy controls in a cross-sectional analysis carried at baseline of an RCT.<sup>20</sup> Higher n-6 DPA also correlated with poorer mental health and less satisfaction with life, while lower n-3 DPA was associated with a higher frequency of self-reported body pain. Data from this small study with multiple statistical testing should be reproduced in larger, longitudinal studies before drawing firm conclusions. In the Three-City cohort study, plasma n-3 DPA was not associated with risk of incident dementia, and n-6 DPA was not available.<sup>21</sup>

### Vitamin D

Beyond its role in bone health, the multiple biological functions of vitamin D are raising increasing interest.<sup>22</sup> In humans, vitamin D

<sup>1</sup>ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, Université de Bordeaux, Bordeaux, France and <sup>2</sup>INSERM, ISPED, Centre INSERM U897-Epidemiologie-Biostatistique, Bordeaux, France. Correspondence: Dr P Barberger-Gateau, Université de Bordeaux, ISPED, Centre de Recherche INSERM U897, 146 rue Léo Saignat, CS61292, Bordeaux, Cedex 33076, France.

E-mail: Pascale.Barberger-Gateau@isped.u-bordeaux2.fr

Received 9 July 2014; accepted 6 August 2014; published online 27 August 2014



comes from both synthesis during exposure to sunlight and food sources, mainly fatty fish as for EPA and DHA. Thus it is difficult to disentangle their respective effects in observational studies. The prevalence of vitamin D deficiency is very high in older adults.<sup>22</sup>

Meta-analyses of case-control studies showed that patients with Alzheimer's disease (AD) had lower serum vitamin D concentrations than controls,<sup>23</sup> with borderline significance after adjusting for age.<sup>24</sup> Longitudinal studies are scarce. Participants in the InCHIANTI cohort study who were severely vitamin D deficient according to their serum 25-hydroxy-vitamin D levels ( $< 25$  nmol/l) had a significantly increased risk of cognitive decline over 6 years.<sup>25</sup> However, this association was not reproduced in another longitudinal study.<sup>26</sup> Several RCTs assessing the impact of vitamin D on cognition, either as primary or a secondary outcome, are in progress (clinicaltrials.gov). In the Women's Health Initiative, supplementation with 400 IU/day of vitamin D, along with calcium, did not result in decreased risk of dementia or cognitive decline, compared with the placebo group over a mean follow-up of 8 years.<sup>27</sup> If ongoing RCTs yield more positive results with higher doses of vitamin D, improvement of cognitive function would be a beneficial side-effect of systematic vitamin D supplementation in deficient individuals.

#### Vitamin K

A growing body of evidence supports a role for vitamin K in brain and cognition.<sup>28</sup> Several vitamin K-dependent proteins contribute to brain function, notably Gas6 which is expressed in the hippocampus of adult rat and contributes to survival of neurons and microglia. Moreover, vitamin K is involved in sphingolipid metabolism.<sup>28</sup> Dietary vitamin K is provided by vegetables, whose consumption is inversely associated with cognitive decline and risk of dementia.<sup>29</sup> Data on the relationship between vitamin K status and cognition in humans are lacking. A single cross-sectional epidemiological study showed a specific relationship between high serum vitamin K (phylloquinone) concentrations and better performances in verbal episodic memory in 320 healthy older adults.<sup>30</sup> These data need to be confirmed in longitudinal studies using serum phylloquinone as a biomarker of vitamin K status.

#### FROM A SINGLE-NUTRIENT APPROACH TO DIETARY PATTERNS

Most previous epidemiological studies and RCTs have focused on single nutrients, ignoring their potential synergistic effects when they are provided in optimal quantities and proportions as in a balanced diet. Considering dietary patterns may lead to a more holistic approach of the diet.<sup>31</sup> Diets rich in fruits, vegetables, vegetable oils, legumes, cereals and fish provide folate, vitamins C and E, carotenoids, polyphenols and long-chain n-3 PUFA along with a low glycemic index that contribute to lower inflammation, oxidative stress and homocysteine concentration, improve vascular status and insulin sensitivity and maintain brain structure and functioning.<sup>32</sup> Moreover, some combinations of nutrients may have synergistic effects that will reinforce their individual properties. For example, antioxidant nutrients can protect long-chain n-3 PUFA from peroxidation to which they are particularly susceptible because of their multiple double bonds.

#### Folate and n-3 PUFA

In the Three-City cohort study, we observed that regular fish consumption was not protective against risk of dementia if not accompanied by regular intake of fruits and vegetables.<sup>33</sup> Accordingly, a placebo-controlled RCT showed that response to n-3 PUFA supplementation (2 g EPA and 1 g DHA per day) for 6 weeks was enhanced in individuals who reported being high consumers of dark-green vegetables.<sup>34</sup> These vegetables are good dietary sources of folate and lutein, a carotenoid, suggesting a

potential interaction between these nutrients and the absorption or bioavailability of n-3 PUFA. However, an another RCT, ancillary study of the SU.FOL.OM3 trial providing B vitamins (including 560 µg/d folate), EPA (400 mg/d) and DHA (200 mg/d) in individuals with a history of cardiovascular disease did not evidence any improvement of cognitive function after 4 years, except maybe in some specific subgroups.<sup>35</sup>

#### Healthy diets

In observational studies, several models of healthy diets combining multiple classes of nutrients have been found to be associated with lower cognitive decline and risk of dementia, the most widely studied being the Mediterranean diet.<sup>2,31</sup> A recent meta-analysis evidenced that higher adherence to a Mediterranean-type diet is associated with a reduced risk of developing MCI and AD.<sup>36</sup> However, the construct of 'Mediterranean diet' may actually encompass very different food and nutrient intakes, depending on the population and geographical area, as most scoring instruments developed to assess adherence to this dietary model are based on sample-specific medians.

Very few RCTs have assessed the impact of shifting to a global healthier diet on cognition. In the PREDIMED-NAVARRA study, participants at high vascular risk randomized to receive intensive advice to adopt a Mediterranean diet enriched with virgin olive oil or nuts scored significantly better than the low-fat control group on the Mini Mental Status Examination (MMSE), a global test of cognitive performance, and the Clock Drawing Test, assessing a wide range of high-level cognitive abilities including executive functions, after 6½-year follow-up.<sup>37</sup> However, cognition was not measured at baseline, and the impact of the diets on cognitive decline could thus not be assessed. The Exercise and Nutritional Interventions for Cardiovascular Health RCT was designed to examine the effects of the Dietary Approach to Stop Hypertension, alone or in combination with weight management through physical exercise and calorie restriction, on blood pressure in overweight adults with high blood pressure. The Dietary Approach to Stop Hypertension diet was associated with a significantly improved psychomotor speed relative to controls.<sup>38</sup>

#### Supplements with multiple nutrients

RCTs with supplements containing various combinations of nutrients trying to reproduce healthy diets can be used to provide a proof of concept of their global impact on cognition. A systematic review of 10 RCTs with multivitamins provided weak evidence of improvement of some cognitive abilities, especially immediate free recall.<sup>39</sup> However, the few available studies were very heterogeneous regarding inclusion criteria, multivitamin constituents and cognitive outcomes. More convincingly, the Souvenir-II placebo-controlled RCT showed that a 24-week supplementation with Souvenaid was able to improve memory in drug-naïve patients with mild AD, that is, an MMSE score  $\geq 20$ .<sup>40</sup> Souvenaid provides EPA, DHA, B vitamins, antioxidants (vitamins C and E, selenium), choline and uridine monophosphate, a specific combination intended to improve synaptic dysfunction in AD. These trials need to be replicated in populations with various degrees of cognitive impairment to identify the most responsive individuals in whom a supplementation could be recommended.

#### BETTER TARGETING POTENTIAL BENEFICIARIES OF NUTRITIONAL INTERVENTIONS

RCTs with nutritional supplements have been carried for the primary or secondary prevention of dementia or for treatment of patients with mild-to-moderate AD (MMSE range 10–26). Some may have failed to show any significant impact on cognition, either because they have included healthy participants who did not decline during the few months of the trial<sup>41</sup> or, conversely,

individuals with AD whose disease was probably too severe to expect a significant effect from nutrition.<sup>42–44</sup> Few studies have specifically targeted the MCI stage.<sup>45,46</sup> Identifying individuals who are the most susceptible to benefit from a nutritional supplementation is therefore of utmost importance.<sup>47</sup>

#### Cognitive criteria and biomarkers of disease progression

In its strict definition, primary prevention takes place before the development of the disease, that is, before any sign of neurodegeneration in the case of AD. Healthy diets in early life could contribute to develop brain reserve.<sup>48</sup> However, the impact of interventions targeting this population on late life cognitive performance is obviously impossible to evidence. Thus primary and secondary prevention of dementia cannot be disentangled. However, cognitive decline is a relatively late event in the course of the disease, and it has been estimated that beta-amyloid, the hallmark of AD, accumulates sharply for about 15 years in the brain before reaching a plateau when the first clinical signs emerge.<sup>49</sup> This 15-year interval would represent a large window for secondary prevention, especially with nutritional interventions. In the absence of specific impairment on most neuropsychological tests at this stage, biomarkers of disease progression could help to identify the best window of opportunity for RCTs with nutritional supplements. These biomarkers include brain atrophy as measured by magnetic resonance imaging, impaired brain glucose metabolism and imaging of beta-amyloid accumulation with positron emission tomography and measurements of beta-amyloid and phosphorylated tau in CSF.<sup>50</sup> However, in our present state of knowledge these should be considered only as research criteria and not used for screening individuals at high risk of dementia.

#### Nutritional criteria

Contrarily to RCTs with drugs, participants in nutritional interventions do not have a null basal level of the nutrient to be evaluated before any supplementation. There is a large day-to-day (intra-individual) variability, especially for nutrients brought by foods that are not consumed on a daily basis such as fish, the main dietary provider of EPA and DHA. In addition, there is a large inter-individual variability, depending on dietary habits. Both sources of variability decrease the power of the study by inflating confidence intervals and may hamper its external validity. Providing individuals who already meet their dietary requirements with additional quantities of nutrients is probably useless and may even be harmful when upper tolerable intake levels are surpassed.<sup>51</sup>

Inclusion criteria should consider nutritional criteria such as low intake of EPA or DHA<sup>52–54</sup> or low blood levels of long-chain n-3 PUFA.<sup>4</sup> Global malnutrition assessed on generic tools such as the Mini-Nutritional Assessment could also help targeting older individuals who could benefit from a nutritional supplementation.<sup>55</sup>

#### Genetic characteristics

Gene × diet interactions on cognition are still poorly understood. The epsilon 4 allele of the Apolipoprotein E (*APOE4*) gene, the main genetic risk factor for AD, is associated with poorer blood response to supplementation with n-3 PUFA<sup>56</sup> or fish intake.<sup>57</sup> Moreover, some epidemiological studies and RCTs have evidenced that *APOE4* carriers had no benefit of n-3 PUFA supplementation or fish consumption on cognition.<sup>58</sup> Other genetic risk factors of AD involved in lipid metabolism could also interact with dietary n-3 PUFA.<sup>59</sup> Further research is needed to better identify these interactions before considering genetic characteristics as inclusion criteria in RCTs with nutritional supplements. In the present state of knowledge, population screening of individuals based on their

genotype would be unethical, but it will probably emerge in the future with the perspective of a 'personalized nutrition'.

#### DESIGNING INNOVATIVE INTERVENTION STUDIES

Inability of most RCTs with nutritional supplements to show any effect on cognitive performance may in part lie in lack of sensitivity of the selected outcomes to change over a relatively short period. Beside traditional criteria based on cognitive and functional decline, biomarkers could contribute to provide evidence of an impact of nutrients and shed new light into their mechanisms of action.

#### Biomarkers of disease progression

Few RCTs with nutritional supplements have used brain or CSF biomarkers of neurodegeneration as outcomes. These biomarkers could be more sensitive and capture early changes in the course of the disease that do not translate yet into improved cognitive performances. In the VITACOG RCT, the rate of brain atrophy over 24 months measured by magnetic resonance imaging as primary outcome was significantly reduced by 30% by supplementation with B vitamins compared with placebo in 168 elderly subjects with MCI.<sup>45</sup> Conversely, there was no effect of DHA supplementation on brain volume change over 18 months in the subsample of 102 participants with AD who had repeated magnetic resonance imaging examination in the Alzheimer's Disease Cooperative Study RCT.<sup>53</sup> However, that subsample size was probably too small to achieve sufficient power.

Forty-nine older adults (20 healthy and 29 with amnesic MCI (aMCI)) were randomized to follow a HIGH or LOW diet for 4 weeks.<sup>46</sup> The HIGH diet was high in fat, especially saturated fat, and with a high glycemic index. Conversely, the LOW diet was low in fat, especially saturated fat, and had a low glycemic index. In aMCI patients, the LOW diet increased beta42-amyloid concentrations in CSF but had the opposite effect in healthy adults, suggesting different mechanisms of action on the accumulation or clearance of beta-amyloid depending on the stage of the disease. The apolipoprotein E concentration in CSF was increased by the LOW diet and decreased by the HIGH diet. These changes in disease biomarkers were accompanied by improvement in delayed visual recall assessed by the Brief Visuospatial Memory Test with the LOW diet.

#### Biomarkers of mechanism of action

Many biomarkers of putative mechanisms of action of nutrients can also be proposed, preferably as secondary outcomes, in RCTs. These include principally measures of inflammation,<sup>60</sup> oxidative stress assessed by isoprostanes in CSF,<sup>44,46</sup> homocysteine<sup>61</sup> and markers of insulin resistance.<sup>46</sup>

In the LOW/HIGH RCT cited above,<sup>46</sup> the CSF insulin concentration increased with the LOW diet in aMCI patients, whereas the HIGH diet lowered the CSF insulin concentration for healthy adults. As an interpretation of these findings, the authors suggest that restoration of normal insulin concentration and activity in central nervous system may have beneficial effects, such as protection against synaptotoxicity of beta-amyloid oligomers. The peripheral metabolic profile of participants was also changed as the HIGH diet increased and the LOW diet decreased plasma lipids and insulin concentration in both groups, indicating improved insulin sensitivity with the LOW diet. The LOW diet also had a favourable impact on oxidative stress with reduced F2-isoprostane concentrations in CSF.<sup>46</sup>

However, some inconsistent results may arise from cognitive and biomarker outcomes. For instance, most RCTs with B vitamins have shown a non-significant impact on cognition<sup>15</sup> despite a strong homocysteine-lowering effect. Paradoxically, supplementation with a cocktail of antioxidant nutrients was able to decrease

oxidative stress as shown by CSF isoprostanes but increased cognitive decline.<sup>44</sup>

The considerable development of omics technologies will give new insight into the potential metabolic pathways involved in the link between nutrition and brain functioning and provide new integrative biomarkers that will help understand their effects. For example, lipidomic profiling recently identified a set of 10 lipids in plasma that predicted short-term conversion to aMCI or AD with a very high accuracy in healthy elderly persons.<sup>62</sup> These phospholipids have essential structural and functional roles in cell membranes, suggesting that their peripheral blood levels could be an early correlate of neurodegeneration in AD. Moreover, this specific lipid profile could be used as inclusion criteria in more efficient RCTs of nutritional interventions with lipids for the prevention of cognitive decline.

## CONCLUSION

Despite the disappointing results of nutritional interventions on cognition so far, there is considerable room for improvement and more evidence-based knowledge on the link between nutrition and cognitive decline in older persons. Trials with nutritional supplements should learn from RCTs with drugs regarding biomarkers of disease progression for inclusion criteria and outcomes. Progress in genetics and omics technologies will also offer new opportunities for research in this field.

## CONFLICT OF INTEREST

Dr Barberger-Gateau reports grants and non-financial support from Danone Research and Vifor Pharma, personal fees and non-financial support from Nutricia, grants and non-financial support from Groupe Lipides et Nutrition and grants from European Union call FP7 HEALTH 2012 INNOVATION1 grant 305483.

## REFERENCES

- Dauncey MJ. New insights into nutrition and cognitive neuroscience. *Proc Nutr Soc* 2009; **68**: 408–415.
- Barberger Gâteau P, Fear C, Samieri C, Letenneur L. Dietary patterns and dementia. In: Yaffe K (ed). *Chronic Medical Disease and Cognitive Aging: Toward a Healthy Body and Brain*. Oxford University Press: New York, NY, USA, 2013; pp 197–224.
- Cunnane SC, Plourde M, Pifferi F, Bégin M, Féart C, Barberger-Gateau P. Fish, docosahexaenoic acid and Alzheimer's disease. *Prog Lipid Res* 2009; **48**: 239–256.
- Dacks PA, Shineman DW, Fillet HM. Current evidence for the clinical use of long-chain polyunsaturated N-3 fatty acids to prevent age-related cognitive decline and Alzheimer's disease. *J Nutr Health Aging* 2013; **17**: 240–251.
- Li FJ, Shen L, Ji HF. Dietary intakes of vitamin E, vitamin C, and beta-carotene and risk of Alzheimer's disease: a meta-analysis. *J Alzheimers Dis* 2012; **31**: 253–258.
- Akbaraly NT, Faure H, Gourlet V, Favier A, Berr C. Plasma carotenoid levels and cognitive performance in an elderly population: results of the EVA study. *J Gerontol A Biol Sci Med Sci* 2007; **62**: 308–316.
- Singh M, Arseneault M, Sanderson T, Murthy V, Ramassamy C. Challenges for research on polyphenols from foods in Alzheimer's disease: bioavailability, metabolism, and cellular and molecular mechanisms. *J Agric Food Chem* 2008; **56**: 4855–4873.
- Clarke R. B-vitamins and prevention of dementia. *Proc Nutr Soc* 2008; **67**: 75–81.
- Crawford MA, Bazinet RP, Sinclair AJ. Fat intake and CNS functioning: ageing and disease. *Ann Nutr Metab* 2009; **55**: 202–228.
- Maruszak A, Pilarski A, Murphy T, Branch N, Thuret S. Hippocampal neurogenesis in Alzheimer's disease: is there a role for dietary modulation? *J Alzheimers Dis* 2014; **38**: 11–38.
- Lim GP, Calon F, Morihara T, Yang F, Teter B, Ubeda O *et al*. A diet enriched with the Omega-3 fatty acid docosahexaenoic acid reduces amyloid burden in an aged Alzheimer mouse model. *J Neurosci* 2005; **25**: 3032–3040.
- Cole GM, Ma Q-L, Frautschy SA. Omega-3 fatty acids and dementia. *Prostaglandins Leukot Essent Fatty Acids* 2009; **81**: 213–221.
- Joseph J, Cole G, Head E, Ingram D. Nutrition, brain aging, and neurodegeneration. *J Neurosci* 2009; **29**: 12795–12801.
- Perez L, Heim L, Sherzai A, Jaceldo-Siegl K. Nutrition and vascular dementia. *J Nutr Health Aging* 2012; **16**: 319–324.
- Wald DS, Kasteurirane A, Simmonds M. Effect of folic acid, with or without other B vitamins, on cognitive decline: meta-analysis of randomized trials. *Am J Med* 2010; **123**: 522–7.e2.
- Mazereeuw G, Lancot KL, Chau SA, Swardfager W, Herrmann N. Effects of omega-3 fatty acids on cognitive performance: a meta-analysis. *Neurobiol Aging* 2012; **33**: 1482.e17–29.
- Dangour AD, Andreeva VA, Sydenham E, Uauy R. Omega 3 fatty acids and cognitive health in older people. *Br J Nutr* 2012; **107**(Suppl 2), S152–S158.
- Miller E, Kaur G, Larsen A, Loh SP, Linderborg K, Weisinger HS *et al*. A short-term n-3 DPA supplementation study in humans. *Eur J Nutr* 2013; **52**: 895–904.
- Guest J, Garg M, Bilgin A, Grant R. Relationship between central and peripheral fatty acids in humans. *Lipids Health Dis* 2013; **12**: 79.
- Milte CM, Sinn N, Street SJ, Buckley JD, Coates AM, Howe PRC. Erythrocyte polyunsaturated fatty acid status, memory, cognition and mood in older adults with mild cognitive impairment and healthy controls. *Prostaglandins Leukot Essent Fatty Acids* 2011; **84**: 153–161.
- Samieri C, Fear C, Letenneur L, Dartigues J-F, Peres K, Auriacombe S *et al*. Low plasma eicosapentaenoic acid and depressive symptomatology are independent predictors of dementia risk. *Am J Clin Nutr* 2008; **88**: 714–721.
- Rosen CJ. Vitamin D insufficiency. *N Engl J Med* 2011; **364**: 248–254.
- Annweiler C, Llewellyn DJ, Beauchet O. Low serum vitamin D concentrations in Alzheimer's disease: a systematic review and meta-analysis. *J Alzheimers Dis* 2013; **33**: 659–674.
- Lopes da Silva S, Vellas B, Elemans S, Luchsinger J, Kamphuis P, Yaffe K *et al*. Plasma nutrient status of patients with Alzheimer's disease: systematic review and meta-analysis. *Alzheimers Dement* 2014; **10**: 485–502.
- Llewellyn DJ, Lang IA, Langa KM, Muniz-Terrera G, Phillips CL, Cherubini A *et al*. Vitamin D and risk of cognitive decline in elderly persons. *Arch Intern Med* 2010; **170**: 1135–1141.
- Slinin Y, Paudel ML, Taylor BC, Fink HA, Ishani A, Canales MT *et al*. 25-Hydroxyvitamin D levels and cognitive performance and decline in elderly men. *Neurology* 2010; **74**: 33–41.
- Rossum RC, Espeland MA, Manson JE, Dysken MW, Johnson KC, Lane DS *et al*. Calcium and vitamin D supplementation and cognitive impairment in the women's health initiative. *J Am Geriatr Soc* 2012; **60**: 2197–2205.
- Ferland G. Vitamin K and brain function. *Semin Thromb Hemost* 2013; **39**: 849–855.
- Loef M, Walach H. Fruit, vegetables and prevention of cognitive decline or dementia: a systematic review of cohort studies. *J Nutr Health Aging* 2012; **16**: 626–630.
- Presse N, Belleville S, Gaudreau P, Greenwood CE, Kergoat MJ, Morais JA *et al*. Vitamin K status and cognitive function in healthy older adults. *Neurobiol Aging* 2013; **34**: 2777–2783.
- Alles B, Samieri C, Fear C, Jutand M, Laurin D, Barberger-Gateau P. Dietary patterns: a novel approach to examine the link between nutrition and cognitive function in older individuals. *Nutr Res Rev* 2012; **25**: 207–222.
- Barberger-Gateau P, Samieri C, Allès B, Féart C. Could nutrition prevent the onset of dementia? Current evidence from epidemiological and intervention studies. *Neurodegener Dis Manag* 2012; **2**: 305–314.
- Barberger-Gateau P, Raffaitin C, Letenneur L, Berr C, Tzourio C, Dartigues JF *et al*. Dietary patterns and risk of dementia: the Three-City cohort study. *Neurology* 2007; **69**: 1921–1930.
- O'Sullivan A, Armstrong P, Schuster GU, Pedersen TL, Allayee H, Stephensen CB *et al*. Habitual diets rich in dark-green vegetables are associated with an increased response to  $\omega$ -3 Fatty acid supplementation in Americans of African ancestry. *J Nutr* 2014; **144**: 123–131.
- Andreeva VA, Kesse-Guyot E, Barberger-Gateau P, Fezeu L, Hercberg S, Galan P. Cognitive function after supplementation with B vitamins and long-chain omega-3 fatty acids: ancillary findings from the SU.FOL.OM3 randomized trial. *Am J Clin Nutr* 2011; **94**: 278–286.
- Singh B, Parsaik AK, Mielke MM, Erwin PJ, Knopman DS, Petersen RC *et al*. Association of mediterranean diet with mild cognitive impairment and Alzheimer's disease: a systematic review and meta-analysis. *J Alzheimers Dis* 2014; **39**: 271–282.
- Martínez-Lapiscina EH, Clavero P, Toledo E, Estruch R, Salas-Salvadó J, San Julián B *et al*. Mediterranean diet improves cognition: the PREDIMED-NAVARRA randomised trial. *J Neurol Neurosurg Psychiatry* 2013; **84**: 1318–1325.
- Smith PJ, Blumenthal JA, Babyak MA, Craighead L, Welsh-Bohmer KA, Browndyke JN *et al*. Effects of the dietary approaches to stop hypertension diet, exercise, and caloric restriction on neurocognition in overweight adults with high blood pressure. *Hypertension* 2010; **55**: 1331–1338.
- Grima NA, Pase MP, Macpherson H, Pipingas A. The effects of multivitamins on cognitive performance: a systematic review and meta-analysis. *J Alzheimers Dis* 2012; **29**: 561–569.



- 40 Scheltens P, Twisk JW, Blesa R, Scarpini E, von Arnim CA, Bongers A *et al.* Efficacy of souvenaid in mild Alzheimer's disease: results from a randomized, controlled trial. *J Alzheimers Dis* 2012; **31**: 225–236.
- 41 Dangour AD, Allen E, Elbourne D, Fasey N, Fletcher AE, Hardy P *et al.* Effect of 2-y n-3 long-chain polyunsaturated fatty acid supplementation on cognitive function in older people: a randomized, double-blind, controlled trial. *Am J Clin Nutr* 2010; **91**: 1725–1732.
- 42 Freund-Levi Y, Eriksdotter-Jonhagen M, Cederholm T, Basun H, Faxen-Irving G, Garlind A *et al.* Omega-3 fatty acid treatment in 174 patients with mild to moderate Alzheimer disease: OmegAD study: a randomized double-blind trial. *Arch Neurol* 2006; **63**: 1402–1408.
- 43 Aisen PS, Schneider LS, Sano M, Diaz-Arrastia R, van Dyck CH, Weiner MF *et al.* High-dose B vitamin supplementation and cognitive decline in Alzheimer disease: a randomized controlled trial. *JAMA* 2008; **300**: 1774–1783.
- 44 Galasko DR, Peskind E, Clark CM, Quinn JF, Ringman JM, Jicha GA *et al.* Antioxidants for Alzheimer disease: a randomized clinical trial with cerebrospinal fluid biomarker measures. *Arch Neurol* 2012; **69**: 836–841.
- 45 Smith AD, Smith SM, de Jager CA, Whitbread P, Johnston C, Agacinski G *et al.* Homocysteine-lowering by B vitamins slows the rate of accelerated brain atrophy in mild cognitive impairment: a randomized controlled trial. *PLoS One* 2010; **5**: e12244.
- 46 Bayer-Carter JL, Green PS, Montine TJ, VanFossen B, Baker LD, Watson GS *et al.* Diet intervention and cerebrospinal fluid biomarkers in amnesic mild cognitive impairment. *Arch Neurol* 2011; **68**: 743–752.
- 47 Dangour AD, Allen E, Richards M, Whitehouse P, Uauy R. Design considerations in long-term intervention studies for the prevention of cognitive decline or dementia. *Nutr Rev* 2010; **68**: S16–S21.
- 48 Benton D. Neurodevelopment and neurodegeneration: are there critical stages for nutritional intervention? *Nutr Rev* 2010; **68** (Suppl 1): S6–10.
- 49 Jack CR Jr., Wiste HJ, Lesnick TG, Weigand SD, Knopman DS, Vemuri P *et al.* Brain beta-amyloid load approaches a plateau. *Neurology* 2013; **80**: 890–896.
- 50 Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger Gateau P, Cummings J *et al.* Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* 2007; **6**: 734–746.
- 51 Bjelakovic G, Nikolova D, Gluud LL, Simonetti RG, Gluud C. Antioxidant supplements for prevention of mortality in healthy participants and patients with various diseases. *Cochrane Database Syst Rev* 2012; **3**: CD007176.
- 52 Yurko-Mauro K, McCarthy D, Rom D, Nelson EB, Ryan AS, Blackwell A *et al.* Beneficial effects of docosahexaenoic acid on cognition in age-related cognitive decline. *Alzheimers Dement* 2010; **6**: 456–464.
- 53 Quinn JF, Raman R, Thomas RG, Yurko-Mauro K, Nelson EB, Van Dyck C *et al.* Docosahexaenoic acid supplementation and cognitive decline in Alzheimer disease: a randomized trial. *JAMA* 2010; **304**: 1903–1911.
- 54 Stonehouse W, Conlon CA, Podd J, Hill SR, Minihane AM, Haskell C *et al.* DHA supplementation improved both memory and reaction time in healthy young adults: a randomized controlled trial. *Am J Clin Nutr* 2013; **97**: 1134–1143.
- 55 Cereda E. Mini nutritional assessment. *Curr Opin Clin Nutr Metab Care* 2012; **15**: 29–41.
- 56 Plourde M, Vohl M-C, Vandal M, Couture P, Lemieux S, Cunnane SC. Plasma n-3 fatty acid response to an n-3 fatty acid supplement is modulated by apoE 4 but not by the common PPAR-alpha L162V polymorphism in men. *Br J Nutr* 2009; **102**: 1121–1124.
- 57 Samieri C, Lorrain S, Buaud B, Vaysse C, Berr C, Peuchant E *et al.* Relationship between diet and plasma long-chain n-3 PUFA in older people: impact of apolipoprotein E genotype. *J Lipid Res* 2013; **54**: 2559–2567.
- 58 Barberger Gateau P, Samieri C, Fear C, Plourde M. Dietary omega 3 polyunsaturated fatty acids and Alzheimer's disease: interaction with apolipoprotein E genotype. *Curr Alzheimer Res* 2011; **8**: 479–491.
- 59 Jones L, Harold D, Williams J. Genetic evidence for the involvement of lipid metabolism in Alzheimer's disease. *Biochim Biophys Acta* 2010; **1801**: 754–761.
- 60 Freund-Levi Y, Hjorth E, Lindberg C, Cederholm T, Faxen-Irving G, Vedin I *et al.* Effects of omega-3 fatty acids on inflammatory markers in cerebrospinal fluid and plasma in Alzheimer's disease: the OmegAD study. *Dement Geriatr Cogn Disord* 2009; **27**: 481–490.
- 61 de Jager CA, Oulhaj A, Jacoby R, Refsum H, Smith AD. Cognitive and clinical outcomes of homocysteine-lowering B-vitamin treatment in mild cognitive impairment: a randomized controlled trial. *Int J Geriatr Psychiatry* 2012; **27**: 592–600.
- 62 Mapstone M, Cheema AK, Fiandaca MS, Zhong X, Mhyre TR, Macarthur LH *et al.* Plasma phospholipids identify antecedent memory impairment in older adults. *Nat Med* 2014; **20**: 415–418.

## NEURODEGENERATION

## Ageing neurons need REST

Individuals who develop Alzheimer's disease typically exhibit neuronal loss in the hippocampus and cortex. In the healthy ageing brain, both of these neuronal populations are preserved, but little is known about the mechanisms that protect neurons against stress and toxic insults. Yankner and colleagues now show that the transcriptional repressor REST (RE1-silencing transcription factor) has a crucial role in neuroprotection during ageing.

The authors carried out a bioinformatic analysis of previous transcriptional profiling studies of the ageing brain, which suggested that REST activation may underlie changes in gene expression during ageing. To explore this possibility, they examined post-mortem samples of the prefrontal cortex (PFC) from young adult and aged individuals. They found that ageing is associated with a significant induction of REST in neuronal nuclei together with increased REST binding to target genes. By contrast, nuclear REST expression was substantially reduced in individuals with mild cognitive impairment (MCI) and almost absent in those with Alzheimer's disease. Furthermore, REST levels closely correlated with cognitive function scores derived from longitudinal neuropsychometric testing.

These findings suggest that deregulation of genes targeted by REST might occur in Alzheimer's disease. To determine which genes are involved, the authors carried out chromatin immunoprecipitation and deep sequencing on a neural cell line and confirmed the results in neuronal nuclei isolated from the human PFC.

This showed that REST regulates many genes associated with cell death pathways and Alzheimer's disease pathogenesis. Analysis of human brain samples demonstrated reduced REST binding and increased expression of many of these genes in individuals with Alzheimer's disease.

The repression of cell death-associated genes suggests that REST might be neuroprotective. Indeed, the authors found that neuronal cultures derived from conditional knockout mice lacking REST were more vulnerable than control cultures to degeneration and cell death induced by oxidative stress or incubation with toxic oligomers of amyloid- $\beta$  (A $\beta$ ). Furthermore, mice lacking REST exhibited progressive neurodegeneration, including neuronal loss in the hippocampus and cortex.

What are the mechanisms by which REST is boosted in the healthy ageing brain? The authors demonstrated the induction of REST in primary cultures of cortical neurons exposed to various stressors associated with ageing. REST was also induced in a neural cell line exposed to the medium in which the 'stressed' neurons had been cultured or to extracts of aged human brain, suggesting that a cell-non-autonomous pathway is involved. Further experiments showed that enhancing signalling through the WNT- $\beta$ -catenin pathway induced REST and that  $\beta$ -catenin levels were increased in



the aged PFC and colocalized in the nucleus with REST, suggesting that this pathway contributes to the induction of REST in the ageing brain.

These findings suggest that loss of neuroprotective nuclear REST contributes to neuronal cell death in Alzheimer's disease, prompting the authors to consider the mechanisms underlying this loss. Dysregulation of autophagy, a process that can sequester proteins in the cytoplasm, occurs in several neurodegenerative diseases, and the authors showed that REST is present in autophagosomes together with misfolded proteins such as A $\beta$ , tau and  $\alpha$ -synuclein in several neurodegenerative disorders. Moreover, activating autophagy in cultured neural cells reduced nuclear REST levels and resulted in REST appearing in autophagosomes.

This study shows that REST induction in the ageing brain is a crucial neuroprotective factor and suggests that boosting this response might provide a strategy to combat age-related neurodegenerative disease, especially Alzheimer's disease.

Katherine Whalley

“nuclear REST expression was substantially reduced in individuals with mild cognitive impairment (MCI) and almost absent in those with Alzheimer's disease”

**ORIGINAL RESEARCH PAPER** Lu, T. et al. REST and stress resistance in ageing and Alzheimer's disease. *Nature* <http://dx.doi.org/10.1038/nature13163> (2014)

# Stress and the brain: individual variability and the inverted-U

Robert M Sapolsky

**It is a truism that the brain influences the body and that peripheral physiology influences the brain. Never is this clearer than during stress, where the subtlest emotions or the most abstract thoughts can initiate stress responses, with consequences throughout the body, and the endocrine transducers of stress alter cognition, affect and behavior. For a fervent materialist, few things in life bring more pleasure than contemplating the neurobiology of stress.**

The early years of the field of stress biology were dominated by the first half of the neuroendocrine loop, namely the ability of the brain to initiate the body's stress response. Stress physiology was born early in the last century, in the Paleolithic era of Walter Cannon implicating the sympathetic nervous system in the 'fight or flight' response, and in Hans Selye identifying glucocorticoids as the other main mediator of the stress response. These foundational findings took the field in many directions. One was the revolutionary work of Geoffrey Harris, Roger Guillemin and Andrew Schally showing that the brain is an endocrine gland, secreting releasing and inhibiting hormones into the hypothalamic-pituitary portal system; in many ways, the decades-long arc of that revolution was bracketed by stress research, in that corticotropin-releasing hormone was the first of the principal hypothalamic hormones whose existence was inferred physiologically and the last to be isolated and biochemically characterized.

The half of the loop concerning the brain regulating the body also encompasses the historic welcoming of psychologists into the field; this came with the demonstration that the stress response, conceptualized in the context of acute physical crisis, can be robustly activated by purely psychological states, such as loss of control, predictability and social support. And that half of the loop also incorporates the fact, first deeply appreciated by Selye, that makes stress biology a branch of medicine: prolonged stress increases the odds of being

sick. This has facilitated the birth of other sub-fields (for example, psychoneuroimmunology), and is now an area of tremendous amounts of reductive research. As a result, we have a fairly good idea as to how, say, a fleeting, stressful thought changes transcriptional events relevant to oxidative metabolism in your big toe.

In many ways, the pendulum has swung and, in recent decades, the field has come to be dominated by the second half of the loop, namely the effects of the peripheral stress response on the brain. This is certainly reflected in the collection of papers in the present issue. Collectively, they highlight a number of important themes.

## Reductive mechanisms underlying stress effects in the brain

A major one, naturally, has been the ever-increasing insights into the mechanisms by which stress affects the brain. For example, decades of work have fruitfully explored the disruptive effects of stress on hippocampal-dependent declarative memory processes and on frontocortical-dependent executive function and behavioral regulation. It is only in more recent years that we have gained insight into the signal transduction pathways and transcriptional events that mediate these stress effects<sup>1,2</sup>. Moreover, some of these mechanistic insights contain surprises that have upended dogma. One, for example, concerns neuroinflammation and the textbook knowledge that glucocorticoids are uniformly anti-inflammatory. However, it is now recognized that this is not always the case, and that glucocorticoids and stress can even worsen facets of neuroinflammation in a brain region-specific manner<sup>3</sup>; the novel depressogenic effects of neuroinflammation<sup>4</sup>

provides a mechanistic route by which stress predisposes to depression.

## Early life stress and adult opportunities that should not be lost

Another theme concerns the complex and supremely important intersection of brain development and adult neuroplasticity. A canonical body of knowledge shows how stress in early life, particularly in the perinatal period, can have predominately adverse neurobiological consequences stretching long into adulthood<sup>5</sup>. These effects can have an extraordinarily long reach, changing the trajectory of brain aging, and even having multi-generational effects, through the non-genetic transmission of behavioral and physiological traits. The mediating mechanism for these long-term effects has increasingly been shown to be epigenetic, a current focus of intense amounts of work.

Running in parallel with this is the evidence of plasticity in the adult nervous system. The excitability of synapses change, dendritic spines come and go within minutes, dendritic processes expand or retract, and circuitry remaps. And then there is, of course, the revolution of adult neurogenesis. Little in the brain, it turns out, is set in stone.

When combined, these two sets of findings produce conclusions that are both salutary and alarming and should galvanize action. First, early life adversity can leave broad and permeating scars of neurobiological dysfunction long into the future, even unto the proverbial generations. Second, there is far more potential for lessening, halting or even reversing these consequences of early life stress in the adult than anyone could have imagined. Third, the longer the intervention is delayed,

Robert M. Sapolsky is in the Departments of Biological Sciences, Neurology and Neurosurgery, Stanford University, Stanford, California, USA.  
e-mail: [sapolsky@stanford.edu](mailto:sapolsky@stanford.edu)



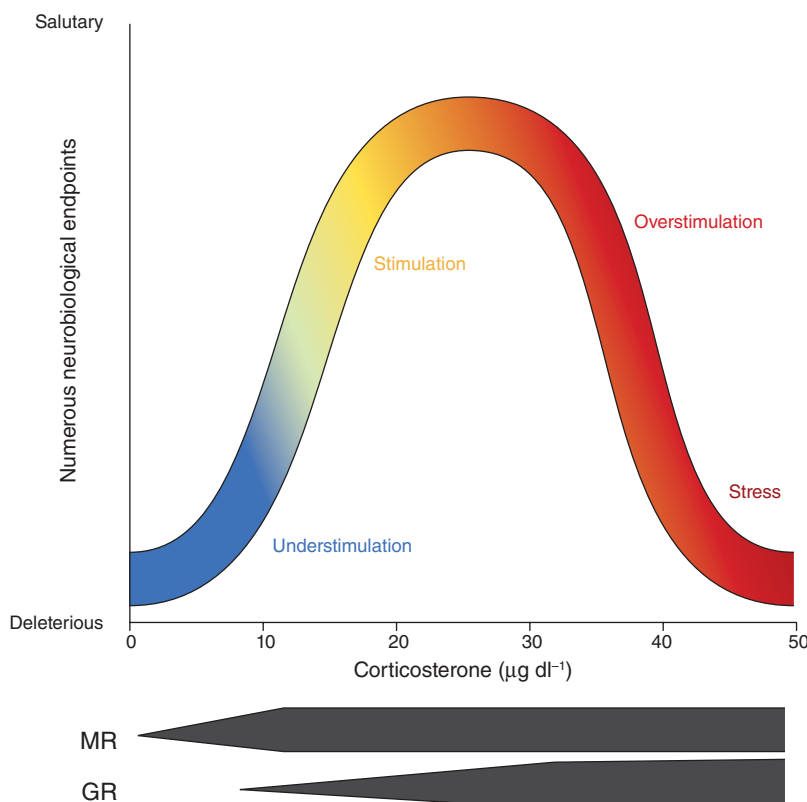
the more of an uphill battle there will be to be make things better.

### Multi-level meanings of stress effects

Another theme is the disconnect between the 'meaning' of a stress effect on the neuron, the brain structure and the organism. A prime example of this concerns the hippocampus, frontal cortex and amygdala<sup>6</sup>. Extensive work has shown that, in the first two structures, stress and glucocorticoid excess will decrease synaptic plasticity, cause atrophy of dendritic processes and even cause a loss of total volume or gray matter volume. These findings have been appropriately interpreted as bad things for the organism, and underlie some of the adverse neurocognitive consequences of depression, post-traumatic stress disorder and childhood poverty. A different picture occurs in the amygdala, particularly the basolateral amygdala, where stress and glucocorticoids increase synaptic plasticity and foster expansion of dendritic processes (which raises challenging questions regarding the mechanisms by which glucocorticoid signaling has diametrically opposite effects in these regions). At first glance, this can readily be interpreted as a good thing—after all, who wouldn't want their neurons to be more strongly and broadly integrated into circuits? However, this amygdaloid hypertrophy can be anything but beneficial, as it contributes to the adverse effects of stress on fear acquisition and extinction: when we are stressed, we learn more readily to be afraid when there is no need to and less readily detect when we are safe. The road to a crippling anxiety disorder is paved with perky amygdaloid synapses.

### The ubiquitous, but nonspecific, role of stress in psychiatric disorders

These amygdaloid effects serve as a segue to the ever stronger evidence for stress as a risk factor for an array of neuropsychiatric disorders, including depression, anxiety, schizophrenia and various addictive behaviors<sup>7,8</sup>. These linkages both take the form of early life stress predisposing toward adult illness and periods of acute stress during adulthood triggering episodes of disease. The breadth of these stress effects implies, at the same time, their nonspecificity. A good example of this is seen with the immunophilin FKBP5, which, as a glucocorticoid receptor co-factor, is highly pertinent to stress neuroendocrinology; variants of the *FKBP5* gene are associated with altered risks of depression, anxiety and PTSD<sup>9</sup>. As another example, consider the *DISC-1* gene (Disrupted in Schizophrenia-1), whose cytoskeletal protein product interacts extensively with the signal transduction pathways of stress signals; despite the specificity implied by *DISC-1*'s name, abnormalities in the structure



**Figure 1** Conceptualization of the inverted-U in the context of the benefits and costs of stress. A broad array of neurobiological endpoints show the same property, which is that stress in the mild-to-moderate range (roughly corresponding to 10–20  $\mu\text{g dl}^{-1}$  of corticosterone, the species-specific glucocorticoid of rats and mice) has beneficial, salutary effects; subjectively, when exposure is transient, we typically experience this range as being stimulatory. In contrast, both the complete absence of stress, or stress that is more severe and/or prolonged than that in the stimulatory range, have deleterious effects on those same neurobiological endpoints. The absence of stress is subjectively experienced as understimulatory by most, whereas the excess is typically experienced as overstimulatory, which segues into 'stressful'. Many of the inverted-U effects of stress in the brain are explained by the dual receptor system for glucocorticoids, where salutary effects are heavily mediated by increasing occupancy of the high-affinity, low-capacity MRs and deleterious effects are mediated by the low-affinity, high-capacity GRs.

or regulation of the protein have been implicated not just in schizophrenia, but in depression and bipolar disorder as well<sup>2</sup>.

Appreciating the importance of stress is critical for understanding the neurogenetics of psychiatric illness, as stress is the poster child for the environment part of gene  $\times$  environment interactions. Beginning with the landmark demonstration of the role of serotonin transporter gene polymorphisms in depression risk<sup>10</sup>, the genetics of mental illness has been repeatedly shown to be about stress/diathesis and about vulnerability to stress.

Thus, a variety of themes appear in these papers. Two even larger ones at least tacitly run through all of them and, I hope, will shape research in stress neurobiology for years to come.

### The humdrum and the fascinating versions of the obligatory question of what is stress

Seemingly within moments of Selye popularizing the word stress in the world of biomedicine,

the definitional debate began. Is stress more about the unpleasantness in the outside world (that is, the stressor) or the resulting changes in the body (that is, the stress response)? Or is it mostly about the neurobiological and psychological space floating between the two? This eventually wearisome debate inevitably constituted the first session of virtually every stress conference for decades; it has finally lost steam, with a sense that the word encompasses all of the above—let a thousand flowers bloom, but just remember to define your particular flower in the Methods section.

The far more interesting version of this question addresses the fact that in any species you'd care to study, different individuals respond to stress differently; there are typically dramatic individual differences as to whether a particular event or internal state is even perceived to be stressful. In other words, what is stress...for this individual? Of course, individual variability is not always the case; a severe injury, a major

burn or a sprint from a predator will reliably activate the stress response and evoke an aversive subjective sense in virtually any organism. But these are not the circumstances of stress that are most pertinent to understanding health and disease in contemporary life. Instead, individual differences are most notable as we navigate life's social exigencies.

Individual differences in stress biology were once mostly an experimental irritant: oh no, because of variability we need a bigger sample size. However, individual variability as to whether something is perceived as stressful, and in resilience and vulnerability to stress-related disease, should be viewed as the most important topic in the field. After all, one person's stress envelope is pushed by getting up at the crack of dawn to bird-watch, and another's by a stint as a mercenary in Somalia. To best appreciate the importance of individual differences in stress responsiveness, it is worth focusing on the single most important concept in the field.

### The inverted-U

When viewed from a distance, the effects of stress on the brain and behavior are often worse than murky, where a stressor might increase some endpoint in one setting, have no effect in another and decrease it in a third. To anyone working in the field, what was apparent from the start was that the response to stress depends on the nature, intensity and duration of a stressor (which at least partially translates into a dependence on the pattern of activation of the sympathetic nervous system, the adrenocortical axis and the other mediators of the stress response).

This represents progress and allows one to begin to corral that array of heterogeneous and often diametrically opposite findings into some groupings: for example, the contrasting responses to physical versus psychological stressors, to biotic versus abiotic stressors, to continuous versus intermittent stressors, and so on. Enormous unifying clarity came with the recognition that, to a large extent, the effects of stress in the brain form a nonlinear 'inverted-U' dose-response curve as a function of stressor severity: the transition from the complete absence of stress to mild stress causes an increase in endpoint X, the transition from mild-to-moderate stress causes endpoint X to plateau and the transition from moderate to more severe stress decreases endpoint X.

A classic example of the inverted-U is seen with the endpoint of synaptic plasticity in the hippocampus, where mild-to-moderate stressors, or exposure to glucocorticoid concentrations in the range evoked by such stressors, enhances primed burst potentiation, whereas more severe stressors or equivalent elevations of glucocorticoid concentrations do the

opposite<sup>11</sup>. This example also demonstrates an elegant mechanism for generating such an inverted-U<sup>12</sup>. Specifically, the hippocampus contains ample quantities of receptors for glucocorticoids. These come in two classes. First, there are the high-affinity low-capacity mineralocorticoid receptors (MRs), which are mostly occupied under basal, non-stress conditions and in which occupancy increases to saturating levels with mild-to-moderate stressors. In contrast, there are the low-affinity, high-capacity glucocorticoid receptors (GRs), which are not substantially occupied until there is major stress-induced glucocorticoid secretion. Critically, it is increased MR occupancy that enhances synaptic plasticity, whereas increased occupancy of GRs impairs it; the inverted-U pattern emerges from these opposing effects.

There are abundant additional examples of inverted-U effects of stress in the hippocampus; this is also the case elsewhere in the brain (mediated by mechanisms other than the MR/GR duo), which is mostly specific to the hippocampus<sup>13–18</sup>.

Thus, a morass of conflicting data is eliminated by recognizing the prevalence of inverted-U's. But even greater insight is provided when considering the collective nature of the various inverted-U's; in general, the effects of mild-to-moderate stress (that is, the left side of the U) are salutary, whereas those of severe stress are the opposite. In other words, it is not the case that stress is bad for you. It is major stress that is bad for you, whereas mild stress is anything but; when it is the optimal amount of stress, we love it.

What constitutes optimal good stress? It occurs in a setting that feels safe; we voluntarily ride a roller coaster knowing that we are risking feeling a bit queasy, but not risking being decapitated. Moreover, good stress is transient; it is not by chance that a roller coaster ride is not 3 days long. And what is mild, transient stress in a benevolent setting? For this we have a variety of terms: arousal, alertness, engagement, play and stimulation (Fig. 1).

Entire careers are spent exploring different parts of the inverted-U. At the far left is the realm of an under-stimulatory environment, with profoundly adverse effects seen in impoverished environments ranging from childhood (with, for example, the nightmarish Romanian orphanages as an extreme) to old age, from humans to zoo animals in sterile cages. The upswing of the inverted-U is the domain of any good educator who intuitively finds the ideal space between a student being bored and being overwhelmed, where challenge is energized by a well-calibrated motivating sense of 'maybe'; after all, it is in the realm of plausible, but not guaranteed, reward that anticipatory bursts of mesolimbic dopamine release are the

greatest<sup>19</sup>. And the downswing of the inverted-U is, of course, the universe of "stress is bad for you". Thus, the ultimate goal of those studying stress is not to 'cure' us of it, but to optimize it.

### Individual differences meet the inverted-U

It is useful to superimpose the world of individual differences in stress responsiveness and vulnerability onto the inverted-U concept, as it allows one to frame the differences in terms of the width, height or symmetry of the U-shaped curve. Most of all, it allows one to hone in on a critical question. For any particular stressor, setting and context, where along the axis of stressor severity is the peak of an individual's inverted-U? In other words, what is the point of transition at which someone's experience turns from stimulation to stress, from striving to learned helplessness, from growing from challenge to crumbling? And from this come additional questions. What are the mechanisms by which adversity shifts inverted-U's to the left (that is, an increased vulnerability to subsequent stress)? What conditions foster inverted-U's that are shifted to the right (that is, resilience)<sup>20</sup>?

These questions should be a primary focus of the field for years to come. It is indisputable that extremes of stress are bad for the brain, a fact pertinent to developmental neuroscience, neurogerontology, and everything in between. It is also equally indisputable that optimal amounts of stress enrich and sustain us. Hopefully, as knowledge continues to accumulate at the rate showcased in this issue, we will gain the means to spend more of our lives experiencing the left sides of our inverted-U's.

### COMPETING FINANCIAL INTERESTS

The author declares no competing financial interests.

- McEwen, B. *et al. Nat. Neurosci.* **18**, 1353–1363 (2015).
- Arntsen, A.F. *Nat. Neurosci.* **18**, 1376–1385 (2015).
- Sorrells, S.F. *et al. Neuron* **64**, 33–39 (2009).
- Hodes, G. *et al. Nat. Neurosci.* **18**, 1386–1393 (2015).
- Tost, H. *et al. Nat. Neurosci.* **18**, 1421–1431 (2015).
- Chatterji, S. *et al. Nat. Neurosci.* **18**, 1364–1375 (2015).
- Holon, N. *et al. Nat. Neurosci.* **18**, 1405–1412 (2015).
- Calhoun, G. & Tye, K. *Nat. Neurosci.* **18**, 1394–1404 (2015).
- Hariri, A.R. & Holmes, A. *Nat. Neurosci.* **18**, 1347–1352 (2015).
- Caspi, A. *et al. Science* **301**, 386–389 (2003).
- Diamond, D.M. *et al. Hippocampus* **2**, 421–430 (1992).
- McEwen, B.S. & Sapolsky, R. *Curr. Opin. Neurobiol.* **5**, 205–216 (1995).
- Chrousos, G. *Nat. Rev. Endocrinol.* **5**, 374–381 (2009).
- Baldi, E. & Bucherelli, C. *Nonlinearity Biol. Toxicol. Med.* **3**, 9–21 (2005).
- Luksys, G. & Sandi, C. *Curr. Opin. Neurobiol.* **21**, 502–508 (2011).
- Prager, E. & Johnson, L. *Sci. Signal.* **2**, re5 (2009).
- Roozendaal, B. *Psychoneuroendocrinology* **25**, 213–238 (2000).
- Sapolsky, R.M., Romero, L. & Munck, A. *Endocr. Rev.* **21**, 55–89 (2000).
- Fiorillo, C.D., Tobler, P. & Schultz, W. *Science* **299**, 1898–1902 (2003).
- Russo, S.J. *et al. Nat. Neurosci.* **15**, 1475–1484 (2012).